Tea for two: language and bilateral trade with China

The article assesses the importance of cultural discourse in economics by exploring the extroversive cultural link between language use frequency and bilateral trade flows. Using linguistic data from Google *n*-grams and data on bilateral trade flows with China over the 1821-2008 period, we test whether the frequency of use of the word 'tea' in a Chinese trading partner's language is associated with the nominal value of its trade flows with China. Our findings suggest that the frequency of use of the word tea predicts current and future trade flows with China, and trade flows affect the frequency of use of the word tea albeit to a lesser extent. The frequency of use of the word tea is influenced by the overall size of the Chinese economy irrespective of the size of the economy of China's trading partner, but smaller countries use the word tea more and increase its use faster. We conclude that the creation of a cultural discourse is endogenous to economic power, and cultural discourse amplifies trade flows. These findings validate the importance of narrative economics and the Culture-Based Development perspective.

Keywords: culture, language, narrative economics, bilateral trade, China

JEL classification: Z12, D02, N70, O43, P25, R12

1 Introduction

Robert Shiller's seminal contribution in the *American Economic Review* (2017) argues that discursive narratives, factual or otherwise, influence the incidence, distribution, and possible control of economic fluctuations. He argues that spoken or written accounts can instinctively and subliminally stimulate emotions and human actions that affect effort levels, purchases, and investments, and thereby shape the evolution of the economy.

Many applications of international trade models estimate the impacts of linguistic distances on trade flows, often using language dummy variables, with linguistic distances claimed to reflect cultural distances. The Sapir-Whorf hypothesis (Whorf, 1956) proclaims that people's perceptions are relative to their spoken language, with an intense debate existing around whether language causally determines or merely reflects post-factum thought and decisions. How language develops and diffuses across space and whether the evolution of language stimulates or is stimulated by international trade is therefore a pertinent question.

The aim of this paper is to assess whether international trade flows affect language use frequency and the proximity of cultural discourse, or whether language use frequency influences trade flows. The findings have substantive value because of the endogeneity problem: there is a lack of clarity on whether preferences and tastes for trading evolve naturally in the process of the economic activity, or whether tastes and preferences are coerced and compelled by a convincing narrative preceding that economic activity. This challenging and pertinent research question has been hitherto inadequately addressed in the literature even though Shiller's contemporary and topical contribution emphasises the prominence of this information. Our study offers a distinctive signal underscoring the relevance of this cultural source of endogeneity bias in the modelling of trade flows.

To achieve this aim, we draw on data relating to the incidence and spread of the Chinese word 'tea' and two of its alternatives ('chai' and 'te'), and assess their associations with trade flows between China and her trading partners. We explore the use frequency of these Chinese words in the recipient trading country's local language and assess whether this incidence predicts future trade flows or whether trade flows predict the subsequent spread and frequency of use of these Chinese words. If past language use frequency is found to predict *current* trade flows, then this corroborates the current practice within the economics literature of using linguistic and cultural distances in gravity models of trade; however, if the results illustrate that the spread of language is associated with *past* economic trade flows then this will point to a clear endogeneity problem and strongly question the use of contemporaneous linguistic and cultural distances in gravity models of trade.

The structure of the paper is as follows. Section 2 reviews recent developments in three fields of economic thought: language and economics, cultural distance and economic flows, and linguistic distance and trade. Special attention focuses on the third of these, as it presents a puzzle regarding the hitherto unclear direction of causality between language and trade. Section 3 presents an empirical Culture-Based Development model for language and bilateral trade. Section 4 outlines the data and estimation strategy. Section 5 presents empirical results and discussion. Section 6 concludes.

2 Language and trade

The debate surrounding whether language shapes the way we think or whether the way we think shapes our use of language – a question known as the Sapir-Whorf hypothesis (see Whorf, 1956) – remains contentious and unresolved. Economic measures, such as population size and firm size (Zipf, 1935, 1949), typically follow the same statistical distribution law as

language (Simon, 1955; Reggiani and Nijkamp, 2015; Modica et al., 2017), and so the question emerges whether there is an inter-temporal causal link between the distribution of linguistic components (words) and the distribution of economic properties of places and entities (Tubadji and Pattitoni, 2022). If there is a statistical link, then it is intuitively appealing to explore the direction of the causal link and hence whether language (and thinking) drives trade or vice versa. Understanding of this causal link would provide crucial insight into the nature of economic processes and help to clarify whether the economy is a product of social and human nature or whether the economic system has an independent and natural law-like existence that influences humans and their behaviour.

In many economic studies, language is considered a proxy for culture (see Guiso et al., 2006) and, in trade models in particular, culture and language are assumed to be constant over time. Both assumptions are problematic because spatial distance (in terms of physical distance or in time) and cultural distance are known to be causally associated (Akerlof, 1997; Inglehart and Wenzel, 2010). When the temporal focus of attention is long enough to include the possibility that culture is able to evolve, then we can appreciate that language also changes in response to existing social and trading interactions, and hence linguistic distances between places change. Parts of cultural evolution are reflected in the evolution of language, including the entry and exit of words some of which may be of foreign origin.

Building on Newtonian ideas of attraction, Ravenstein (1885) explored propensities to migrate between two geographical areas and Isard (1954) formalised this idea into a gravity model[1]. The incorporation of variables that reflect cultural and linguistic distances into gravity models enhances the statistical power to predict economic and human entities. This applies when the dependent variables in gravity models include financial flows (Hahn, 2014), migration (Bratti et al., 2014), tourist flows (Redondo-Carretero et al., 2017; Lien et al., 2017), traffic, mail flows, telephone calls, remittances (Lueth and Ruiz-Arranz, 2006) and, of course, trade (Hutchinson, 2005; Dunlevy, 2006). A population with a particular cultural background will then lead to greater number of tourists, migrants, and/or trade from certain culturally close background (Tubadji and Nijkamp 2015; Tubadji and Nijkamp 2018). Our proposition here is that the cultural impact may have a different direction when linguistic proximity background is concerned.

In his seminal work on narrative economics, Shiller (2017) emphasizes that economic phenomena are shaped by the spread of narratives, i.e. the words that describe a certain viewpoint or a logical sequence, where the spread of narratives is conceptualized in a similar way to the spread of infectious diseases. Naturally, the spread of a narrative will be associated with the spread of the frequency of the use of words that make up that narrative, with Tubadji et al. (2022) showing that word frequency has associations with Zipf's and Gibrat's laws.

Linder (1961) offered an alternative perspective to the standard gravity model of bilateral trade by asserting that patterns of trade respond to aggregate preferences for goods within countries, where those countries with more similar preferences for goods develop

[1] Mathematical formalization of the gravity model was offered by Anderson (1979), increasing the precision of understanding of the degree of bias in efficiency in the gravity equation related estimations and their sensitivity to the economic structure of the markets of the trading countries. As noted by Anderson and van Wincoop (2003) the gravity model has demonstrated that it easily fits the data, but the improved precision of the Anderson (1979) approach allows more precise estimation of the effects from the determinants in the gravity equation. Hence, when gravity equations are analyzed, unless the authors are overconfident in their data and modelling, the safest source for analysis is the statistical significance and not the size of the coefficients. We will stick to this conservative approach here.

more similar industrial bases and subsequently trade more. But the question remains why similar preferences emerged and whether this is due to a shared narrative.

Here we suggest one quantitative and one conceptual causal proposition for the economic analysis of language. There is a difference between linguistic distance, that is based on language characteristics traits, and linguistic frequency. We suggest from a quantitative perspective that linguistic frequency is superior to linguistic distance for the approximation of cultural distance. We expect linguistic distances to evolve over time due to changes in the frequency of use of foreign words. Hence, changes in the frequency of use of specific words will more accurately reflect changes in cultural distances than will the use of static indices of linguistic distances that compare languages at a fixed point in time.

The rate of diffusion of foreign words stems from both the rate and duration of human migration and from the spread of knowledge about foreign cultures and languages driven by intellectual curiosity. However, it could be a result of economic exchanges between two countries, such as the incidence and volume of trade between them. This is particularly relevant for the spread of foreign words associated with imports that did not previously have local equivalent names or translations, and in these circumstances the local inhabitants are likely to adopt and use linguistic terms originating from the exporting country.

Tea is a prime example of a product that did not have a local name prior to its arrival in overseas territories.[2] It is believed that tea (i.e., the product) originated in China[3] and then began to be exported across the globe. Interestingly, however, this drink adopted two different names depending on the way that the product was distributed through evolutionary trade networks: if the product was transported through seafaring trade routes, then it became known as *tea* (or *te*). Alternatively, if the product was transported over land through Central Asia and on through Russia then the product became known as *chai* (Saberi, 2010; World Atlas of Language Structures, 2019).

Tea is an important part of cultural evolution, not least in the UK where it remains part of the cultural identity. Although tea was introduced in the UK around the 1660s and became part of the staple diet of all UK social classes within the 18th century and a necessity by the 1890s (Gazeley, 1984, p.318), the British activity of pausing for afternoon tea only became a social event during the 1880s when aristocratic women would dress in long gowns, gloves and hats to be seen to drink tea in drawing rooms (Johnson, n.d.). In the 19th century, tea became the preferred drink rather than coffee, because tea tasted good even when it was diluted, which is often how the poor consumed it to save money (Mintz, 1985). Further, Johnson (1988, p.37) records that "advertising directed at the working-class readers of the *News of the World* (a now defunct national newspaper) at the turn of the century focused on selling branded tea, cocoa, and soap, not on selling patterned plates or china ornaments." An argument can be made that tea was instrumental not only in cultural change, but also in political change, with Chrystal (2014) arguing that the rise of fashionable tea rooms provided a safe place for women to meet and strategize about political campaigns, such as the struggle for the vote.

Tea entered the social narrative and the popular phrase "not for all the tea in China" originated around the late 19th / early 20th century and derives from the knowledge that China was well-known for producing tea in huge quantities. The question then is whether the narrative spread of the word tea (or *te* or *chai*) in a trade recipient country led to greater imports of tea and other products from China, whether greater imports from China intensified

[2] Tea is the world most popular non-alcoholic drink (MacFarlane and MacFarlane, 2004) and Turkey's population consumes the most amount of tea per capita (Euromonitor International, 2013).

[3] There are no known wild populations of the tea plant, so it is not possible to be precise about its original native location (Meegahakumbura *et al*., 2016).

the spread and use of this word in a recipient country, and/or whether this association is contemporaneous. To the knowledge of the authors, this assessment of the direction of causality between language frequency and bilateral trade flows has not been examined before in the economics literature and yet it is crucially important because it leads to conclusions about the importance of narratives in the shaping of trade and economic development.

3 A Culture-Based Development (CBD) assessment of language frequency and trade

In order to assess the relationship between linguistic distance and trade flows, we build on the established Culture-Based Development (CBD) approach (Tubadji, 2013, 2013) by modelling bilateral trade as:

$$TRADE\_FLOW_{ij} = f(GDP_i, GDP_j, Distance_{ij}, WORD\_J_i) \quad (1)$$

where *i* is the recipient country, *j* is the country of origin (i.e. China), $TRADE\_FLOW_{ij}$ *is the* amount of trade between China and a trade partner, GDP_i is the GDP in the partner country, GDP_j is the GDP in the origin country, $Distance_{ij}$ is a vector of relevant types of physical distances (by road, sea) with their squared terms representing the nonlinear distances between the two countries (reflecting travel costs and a distance decay effect), and $WORD\_J_i$ is the frequency of use of the country of origin's linguistic components that are in use in the trade partner's language (reflecting the power of the narrative about China in the recipient country).

Component $WORD\_J_i$ is of special interest because it captures the impact of the spread and intensity of use of Chinese words on bilateral trade flows between China and the recipient country. If Chinese words reflect the dispersion of narratives regarding China that build a China-friendly discourse reflected in fashion and in word of mouth which people subsequently engage in the consumption of the associated item, then economic activity is driven by narrative and culture. Alternatively, if the value of bilateral trade of a particular good subsequently introduces or changes the frequency of use of the corresponding Chinese word(s) in the recipient country's language, then trade influences culture and narratives. Hence, application of model (1) over time will be plagued with endogeneity. Our empirical analysis will focus on establishing whether there is such an endogenous dependency between language and trade flows, and hence we are ultimately interested in the performance of model (2), such that:

$$WORD\_J_{i(n)} = f(TRADE\_FLOW_{ij(n-k),}) \quad (2)$$

where word frequency *WORD_J* is measured in time *n* and trade flows now correspond to their lagged values (*n*-*k*) and could be quantified as a function of GDP_i, GDP_j, and $Distance_{ij}$. In our estimations, we assume that these models are represented by Cobb-Douglas-type functions and can therefore be transformed into linear logarithmic form. Hence, our estimations take the natural logarithm values of the determinants above described.

4 Data and Method

Data were drawn from two main sources. Linguistic data were extracted from Google *n*-grams. The Google *n*-grams relies on a corpus of digitized texts representing over 4% sample of all books stored in the world libraries over the last 200 years. Its records are available on words as lowest level of observation (although combinations of words and phrases are also possible to search for) and are provided on yearly basis[4]. The countries available in the Google *n*-grams dataset are Germany, Spain, Italy, Russia, France, the UK, and the US. This data corresponds to the 1821-2008 period and specifically records the frequency of use of the

[4] See Michel et al. (2011) and Lin et al. (2012) for more details and early applications in non-economic analysis.

word ‘tea’ in a country that is a trading partner of Chinese.[5] This is motivated by the fact that the word tea and its variants are of Chinese origin. Data on nominal bilateral trade flows with China were obtained from the Centre d’Etudes Prospectives et d’Informations Internationales (see Fouquin and Hugot, 2016) and limited so that they correspond to the same time period as the linguistic data. It is convention for gravity models to be parameterized using nominal values and we selected for analysis the bilateral trade flows between China and respectively trade partners for which we have Google *n*-grams data: Germany, Spain, Italy, Russia, France, the UK, and the US[6]. Descriptive statistics for all variables are available in Appendix 1.

Diagrams representing the evolutionary trajectory of the word tea (and its variants) for each country are presented in Figure 1(a-f). The sudden increase in the line for Germany is intriguing. The political isolation of Germany by the Anglo-Japanese Alliance stimulated the German-Chinese-American unification into an informal alliance in 1907, and this may have enhanced recognition of Chinese tea products.

{Figure 1(a-f)}

Figure 2(a - g) displays our data on total trade flows. Estimations of gravity models typically parameterise the dependant variable with the sum of inflows and outflows of trade between two countries. As can be seen from Figure 2, there appears to be a general positive association between total trade flows and the frequency of use of the word tea across all countries over the time period. However, the data appears to segment into clusters, so we explored the variation by country and found that while there is a positive association for most countries the association for Great Britain appears to be the reverse, albeit with a strong co-dependence. This corroborates nuances found in the literature that there are differences between the effects of trade inflows and outflows, and hence we will explore the separate effects between the frequency of the word tea and inflows, outflows, and total trade flows.

{Figure 2(a-g)}

We have a series of relevant controls in our gravity models, such as distance to and contiguity with China. We use two different distance measures: road distance and sea distance, both in kilometres. This is required since we know that some countries trade goods overland while others trade via sea routes, which is an important dichotomy associated with the linguistic marker for tea in the recipient language (as mentioned above, *te* was used if transport was via sea and *chai* was used if transported overland (Saberi, 2010; World Atlas of Language Structures, 2019)). We use both distances and their square values to control for the decay of distance known to be present in gravity models (Fouquin and Hugot, 2016[7]). Our

[5] Portugal ran Macao until 1999, but we are unable to analyse the Portuguese language because Google libraries do not provide *n*-grams for this language at the current time.

[6] While the intensity of the use of the word ‘tea’ might have become seemingly very high across the world, we still expect that the initial conditions of language penetration from early periods contribute to preserving potentially significant differences in the intensity of the use of the word across space nowadays. This is the empirical question that we are set to explore.

[7] Fouquin and Hugot (2016, p.12) show that total nominal imports are greater than total nominal exports for most years and they argue that this is due to two main reasons. First, customs conventions specify that trade is reported Free on Board to the exporting country and includes the Cost of Insurance and Freight at destination. Second, importing countries have greater incentives to carefully report in-flowing trade as governments generally levy taxes on imports. They also note that this gap reduces from being greater than 10

results do not include a control for the colonial past because there is no variation in this indicator (obtained from CEPPI along with the rest of our controls) for the countries in our dataset.

We test our CBD models, (1) and (2), by first establishing whether the frequency-of-use-of-language variable – which is our quantitative measure of cultural bias and which is consistent with narrative economics – has explanatory power in the standard gravity model. Then we assess the evidence for any reverse causality between bilateral trade flows and language use frequency.

In the first step of our analysis we test model (1) using a pooled cross section with year and country fixed effects. We also verify our result by using a panel approach with fixed effects.[8] We re-estimate operationalizations of model (1) using trade inflows, outflows, and then total trade flows.

In the second step, we test model (2) where the dependent variable is the frequency of use of the word tea (and its variants) and the main explanatory variable of interest is the value of bilateral trade flows between China and a trading partner. We assess the associations between the three alternate indicators of trade flows and the standard gravity model determinants: GDP values for China and for the respective trading partner. Finally, in a third step, we re-estimate models (1) and (2) with the inclusion of lagged values of their determinants to cross-check the causal effect between language and trade flows.[9]

All panel estimations were estimated using fixed effects. The Breusch and Pagan test was used to establish whether pooled OLS is significantly different than a random effects panel estimator. All tests pointed in favour of the panel estimator. A Hausman test was also estimated and found to be statistically significant, as seen from Appendix 2, and therefore the preferred estimator has fixed effects.

Ultimately two robustness checks were introduced. The first one examines the effect of different cut off periods in the datset, namely we keep the observations from different starting pints, such as since 1900, 1925, 1950, 1975 and seek to identify whether the exclusion of earlier observations affects the results. The second robustness check examines whether the omission of Russia and Germany will affect the results, As seen from Figure 3 below, these countries have respectively a very low number of observations (Russia) or start in a period later than the other countries (Germany).

{Figure 3}

5 Results

Implementation of our strategy enables an assessment of the association between the frequency of use of the word tea and bilateral trade flows between China and a selection of her partner countries. Table 1 reveals that the frequency of use of the word tea is strongly positively associated with trade flows both to (specification 1) and from (specification 2)

---

percent in the nineteenth century, to less than 5 percent after World War II, and to almost zero in the modern era.

[8] In the panel setting, we must drop the physical distance variable, as it is constant over time. The panel estimator considers the clustering over time and space, hence we have to drop the use of fixed effects for time and space here as well.

[9] All results were recalculated with controls for tariffs in the partner country. The other results remain substantively the same and the tariffs variable had the expected negative sign. However, the tariffs variable was not always statistically significant and it decreases the degrees of freedom, so we present what we believe is a parsimonious model.

China. The economic size of a country seems to determine trade inflows to China, and the size of China determines the size of its outflows. The other gravity model controls are all statistically significant. The overall explanatory power of the pooled OLS models in Table 1 is remarkably good but note that certain differences become more pronounced when we use panel fixed effect estimators. First, the frequency of use of language seems to predict only the inflows to China but the outflow from China is merely a function of its economic size in terms of GDP. This finding suggests that increasing cultural proximity between China and her trading partners affects the willingness of China to import from them. These results underscore the importance of narrative economics by not only asserting that the frequency of use of the word tea reflects the growing demand for tea in the staple diet and hence greater imports, but also by implying that the expanding importance of tea in the evolving national culture may instinctively and subliminally stimulate emotion and human action that enhances purchases in China and thereby shape the evolution of imports from a trading partner.

We re-estimated model (1) using the total volume of trade between China and its partners as a dependent variable using pooled OLS and panel fixed effects, and present these results in Table 3. These results corroborate the view that, despite differences between the effects for inflows and outflows, we can continue to claim that in general the frequency of the use of the word tea seems to be a strong significant predictor of trade flows between China and her trading partners.

{Tables 1 – 3}

Next we assess the reverse causal link from bilateral trade on to the frequency of use of the word tea. Table 4 presents our pooled cross section estimations for the effect of past trade inflows and outflows on the frequency of use of the word tea; we use both the direct quantification of the trade flows and their indirect quantification through their gravitational components (GDP values of importers and exporters, and distance and decay of distance factors for trade), as presented in specifications 1-4. Results support the view that bilateral trade flows affect the use of the word tea, and it appears that both GDP and distance matter for trade flows and the exotic flavour of China for its partners. Specifically, decay of the distance effect is observable, which is consistent with the idea that while initially attraction decreases with greater distance due to associated economic costs, much greater distances create an exotic attraction that increase in importance the further a partner is located from China. We extend the analysis to a panel data estimator with fixed effects, as presented in Table 5, and this more powerful estimator reveals a strong significant effect from trade inflows and outflows to the frequency of the word tea in a recipient country. This more precise estimation shows that the rest of the components of the gravity model (except for the GDP of the partner country) generally lose their importance in determining the frequency of the use of the word tea, but the trade inflows and outflows remain strong predictors of the use of the word in the panel fixed effects estimator. Table 6 confirms that we can generalize the effect of total trade flows on language. This indicates that we indeed have an endogeneity of language, and it should be considered seriously when linguistic distance is employed in gravity models.

{Tables 4 – 6}

Finally, to further clarify the causal direction between language and trade flows, we assess the importance of lagged values of the two variables over each other's evolution. Using the most reliable panel fixed effect estimator, Table 7 reveals strong significance of the lagged effect from the size of the sending partner for trade inflows into China, and these

effects carry over, albeit less prominently, when the total trade flows variable is used as a dependent variable (specification 3). Language influences direct trade inflows, although it does not have a lagged influence and is not confirmed as a factor determining total trade flows. Table 8 reveals that the lagged effect of total trade flows on language use frequency is the only effect from trade to language when lagged variables are used. As seen from Specification (3) in Table 8, lagged values of total trade flows is a strong predictor for the future frequencies of the use of the word tea. This supports our concerns about the reverse causality of trade flows and language frequency, which demonstrates that there is a strong possibility that the evolution of language is endogenous to gravity models over the long run, and hence questions their use as supposedly exogenous variables in gravity models.

{Tables 7 and 8}

Future estimations of gravity models should reflect on whether linguistic components are exogenous. As language is used as a standard proxy for culture in gravity models, such models need to account for this culturally evolving narrative economics component if they are to be correctly specified and unbiased. Further research should explore the possibility that colonies and spatial contiguities permit faster and deeper integration of language and expedite the process of cultural convergence.

Finally, our robustness checks, using different time cut off periods and excluding the countries Germany and Russia from our estimations, show that the results are particularly robust. Appendix 3 and Appendix 4 show the results respectively for time cut-off periods and results without Germany and Russia. The results become affected only in columns 4, when the number of observations in our analysis falls too much because of the intentional omission of observations, required by the robustness check designs.

6 Conclusion

This article makes a novel contribution to the literature as it explores the importance of the narrative approach to economics through the proliferation of language. It relates this idea to the famous question about the Zipf law distribution of words in language, which is a distribution found among many economic processes (Tubadji and Pattitoni, 2022). Our study makes two novel contributions. First, it demonstrates that linguistic distance is not a constant but a changing variable over time, and second, it explores the direction of causality between language and trade flows using the standard gravity model. We carry out this research by exploiting relatively new data from Google on *n*-grams and specifically the frequency of the use of the word ‘tea’ and its variants over time.

An important finding is that the linguistic factor is actually endogenous to economic trade and is strongly influenced by the economic size of the exporting country. The latter is especially true for lagged values of trade flows, and hence we reveal a reverse causality between trade flows and language that poses an important question on the modelling of linguistic distance in gravity models. Our findings suggest the presence of endogeneity problems in gravity models over time periods that include the possibility of cultural change that do not account for the reverse effect of trade on linguistic distance.

The extant literature reveals that there are important within group ethno-linguistic variations (Gören, 2017). We demonstrate that these variations, and perhaps all ethno-linguistic identities, change over time in a manner that is dependent on economic processes. Our findings show that even when we try to account for differences between inflows and

outflows, and find such differences in terms of the impact of language on trade, when we consider the effect of trade on language we observe a seemingly persevering effect from total trade flow to language frequency. We also find a distance decay effect on language, which seems to corroborate an association between distance and expectations of tastes for the exotic. While language use initially decreases with distance (associated with the lower use of goods due price increasing transportation costs), after a turning point greater distances contribute to increases in cultural attraction, reflected in the increased frequency of use of foreign words. Thus, our study is an important signal that while it is generally accepted that culture affects morality and thus economic development (Lee and Bohanon, 2019), it seems that culture is endogenous to economic interests and by extension morality is also likely to be driven by the economic process. This important link should be account for in future studies to avoid biases from the endogeneity of culture and from threats for cultural self-righteousness. The latter supports important philosophers from Ludwig Wittgenstein to Michel Foucault who asserted that the cultural perception of reality is constructed and deconstructed, and hence evolving and non-deterministic in nature. We add to this that the cultural perception of reality is sensitive to influences from economic interests themselves. If our findings hold true in such inter-group contexts, this might have important implications for minorities which are culturally and economically supressed.

Furthermore, our results are generalizable to culture per se. Since language is a proxy for culture, the cultural distance between countries is also changing due to the influence of economic transactions, suggesting that the cultural factor needs to be accounted for in both its traditional and modern components (i.e., both cultural heritage and the living culture, in the terminology of Culture-Based Development if the model is to be fully culturally identified**.**

References

Akerlof, G. (1997) 'Social distance and social decisions,' *Econometrica*, 65(5), 1005-1027.

Anderson, James, E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1), 106-116.

Anderson, James, E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *The American Economic Review*, 93 (1): 170-192.

Bratti, M., L. De Benedictis and G. Santoni (2014) 'On the pro-trade effects of immigrants,' *Review of World Economics*, 150(3), 557-594.

Chrystal, P. (2014) *Tea: a very British beverage*, Amberley Publishing Limited

Dunlevy, J. (2006) 'The influence of corruption and language on the pro-trade effect of immigrants: evidence from the American States,' *Review of Economics and Statistics*, 88(1), 182-186.

Euromonitor International (2013) 'Turkey: second biggest tea market in the world', *Market Research World*, 13/May/2013.

Fouquin, M. and J. Hugot (2016) 'Two centuries of bilateral trade and gravity data: 1827-2014,' CEPII working paper, #2016-14.

Gazeley, I. (1984) 'The standard of living of the working classes, 1881-1912: the cost of living and the analysis of family budgets,' *Oxford University D.Phil. Thesis*.

Gören, E. (2017) Consequences of Linguistic Distance for Economic Growth, *Oxford Bulletin of Economics and Statistics*, 80(3): 625–658.

Guiso L, P. Sapienza and L. Zingales (2009) 'Cultural biases in economic exchange?' Quarterly Journal of Economics, 124(8), 1095-1131.

Guiso, L., P. Sapienza and L. Zingales (2006) 'Does culture affect economic outcomes?' *Journal of Economic Perspectives*, 20(2), 23-48.

Hahn, F. (2014) 'Culture, geography and institutions: empirical evidence from small-scale banking,' *Economic Journal*, 124, 859-886.

Hutchinson, W. (2005) "Linguistic distance" as a determinant of bilateral trade, *Southern Economic Journal*, 72(1): 1-15.

Inglehart, R. and C. Welzel (2010) Changing Mass Priorities: The Link Between Modernization and Democracy, *Perspectives on Politics*, 8(2): 551-567.

Isard, W. (1954) 'Location theory and trade theory: short-run analysis,' *Quarterly Journal of Economics*, 68(2), 305-322

Johnson, B. (no date) 'Afternoon tea,' *Historic UK: the history and heritage accommodation guide*, downloaded from https://www.historic-uk.com/CultureUK/Afternoon-Tea/ on 20/8/2019

Lee, D. and C. Bohanon (2019) Economics and novels: good, evil and becoming better people, *Journal of Cultural Economics*, 43:527–544.

Lien, D., F. Yao and F. Zhang (2017) Confucius Institute's effects on international travel to China: do cultural difference or institutional quality matter? *Applied Economics*, 49(36): 3669-3683.

Lin, Y., J. Michel, E. Aiden, J. Orwant, W. Brockman and S. Petrov (2012) Syntactic Annotations for the Google Books Ngram Corpus. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* Volume 2: Demo Papers (ACL '12) (2012).

Linder, S. B. (1961) *An essay on trade and transformation*, New York: Wiley & Sons

Lueth, E. and M. Ruiz-Arranz (2006) 'A gravity model of workers' remittances,' *IMF working paper*, WP/06/290

MacFarlane, A. and MacFarlane, I. (2004) *Green gold: the empire of tea*, Ebury Press

Meegahakumbura, M. K., Wambulwa, M. C., Thapa, K. K., Li, M. M., Möller, M., Xu, J. C., Yang, J. B., Liu, B. Y., Ranjitkar, S. and Liu, J. (2016) 'Indications for three independent domestication events for the tea plant and new insights into the origin of tea germplasm in China and India revealed by nuclear microsatallites', *PLoS One*, 11(5), e0155369.

Michel, J., Y. Shen, A. Aiden, A. Veres, M. Gray, W. Brockman, The Google Books Team, J. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. Nowak, and E. Aiden (2011) Quantitative Analysis of Culture Using Millions of Digitized Books, *Science*, 331(6014): 176-182.

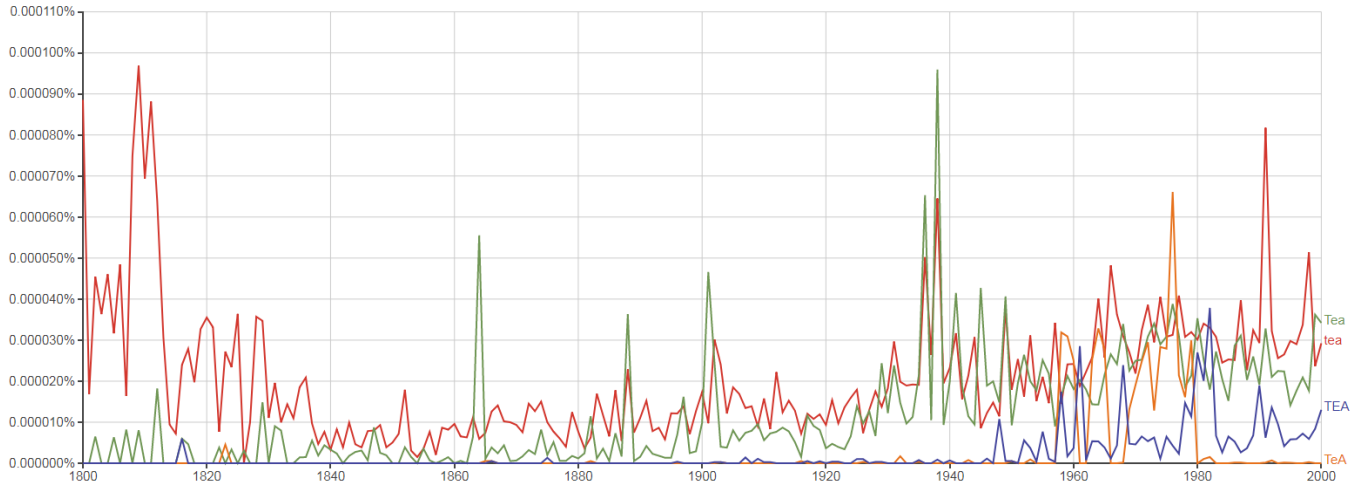
Mintz, S. W. (1985) *Sweetness and power*, Penguin Books: New York

Modica, M., A. Reggiani and P. Nijkamp (2017) 'Methodological advances in Gibrat's and Zipf's laws: a comparative empirical study on the evolution of urban systems,' in H. Shibusawa et al. (eds.) *Socioeconomic Environmental Policies and Evaluations in Regional Science*, 37-59

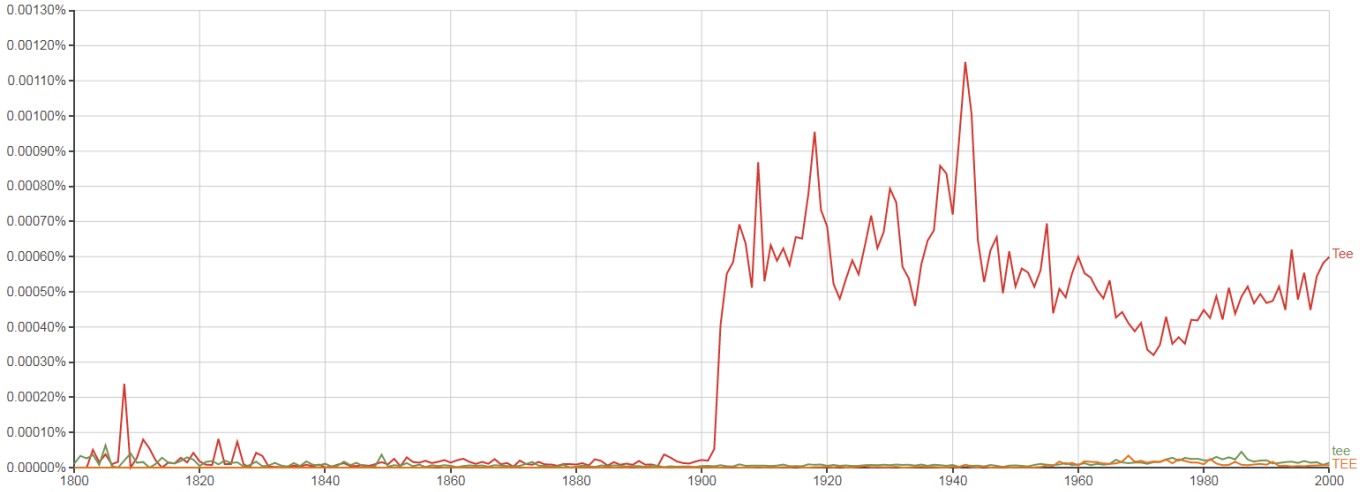
Ravenstein, E. (1885) 'The laws of migration,' *Journal of the Statistical Society of London*, 48, 167-235

Redondo-Carretero, M., C. Camarero-Izquierdo, A. Gutiérrez-Arranz and J. Rodríguez-Pinto (2017) Language tourism destinations: a case study of motivations, perceived value and tourists' expenditure, *Journal of Cultural Economics*, 41(2): 155-172.
Reggiani, A. and P. Nijkamp (2015) Did Zipf Anticipate Spatial Connectivity Structures?, *Environment and Planning B: Urban Analytics and City Science*, 42: 468–489.
Saberi, H. (2010) Tea: A Global History, London: Reaktion Books.
Shiller, R. (2017) Narrative Economics, *American Economic Review* 107(4): 967-1004.
Simon, H. (1955) "On a class of skew distribution functions," *Biometrika*, 42: 425–440.
Tubadji, A. (2013) Culture-Based Development: Culture and Institutions – Economic Development in the Regions of Europe. *International Journal of Society Systems Science*, 5(4):355-391.
Tubadji, A. (2012) Culture-Based Development: Empirical Evidence for Germany. *International Journal of Social Economics*, 39(9):690-703.
Tubadji, A. and P. Pattitoni (2022) Language and the Regional Economy: Cultural Persistence, Resilience or Path-Dependence? Manuscript.
Tubadji, A. and P. Nijkamp (2018) 'Revisiting the Balassa–Samuelson effect: International tourism and cultural proximity,' *Tourism Economics*, 24(8), 915-944.
Tubadji, A. and P. Nijkamp (2015) Cultural Gravity Effects among Migrants: A Comparative Analysis of the EU15, *Economic Geography*, 91(3): 344-380.
Whorf, B. (1956) Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. Edited by John B. Carroll. Cambridge, MA: MIT Press.
Zipf, G. (1935) *The psychology of language*, Houghton-Mifflin.
Zipf, G. (1949) *Human behavior and the principle of least effort*, Addison-Wesley, Cambridge, MA

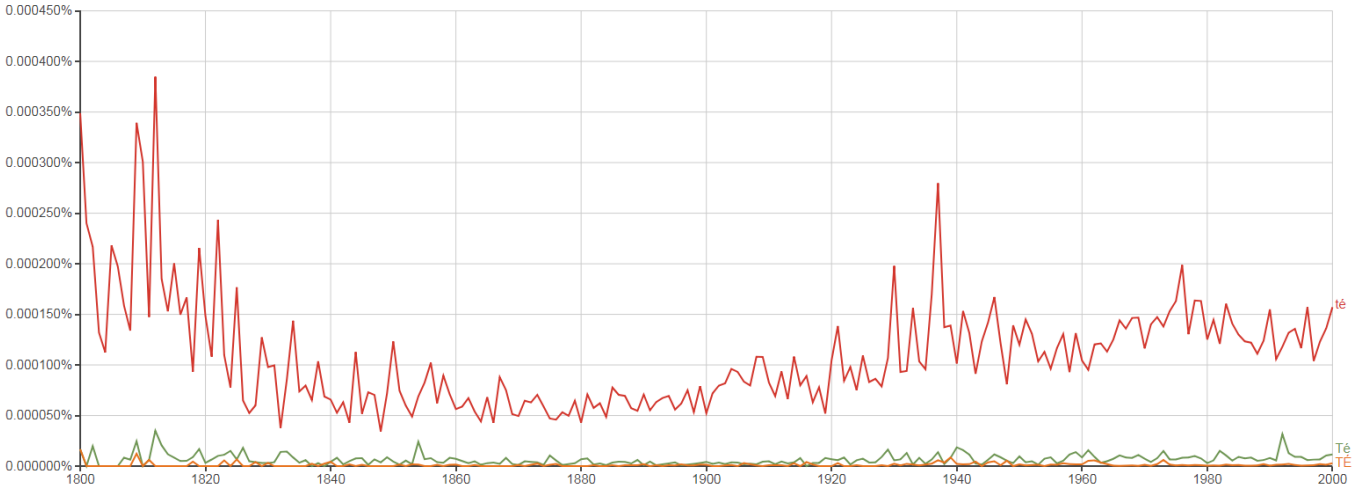
French



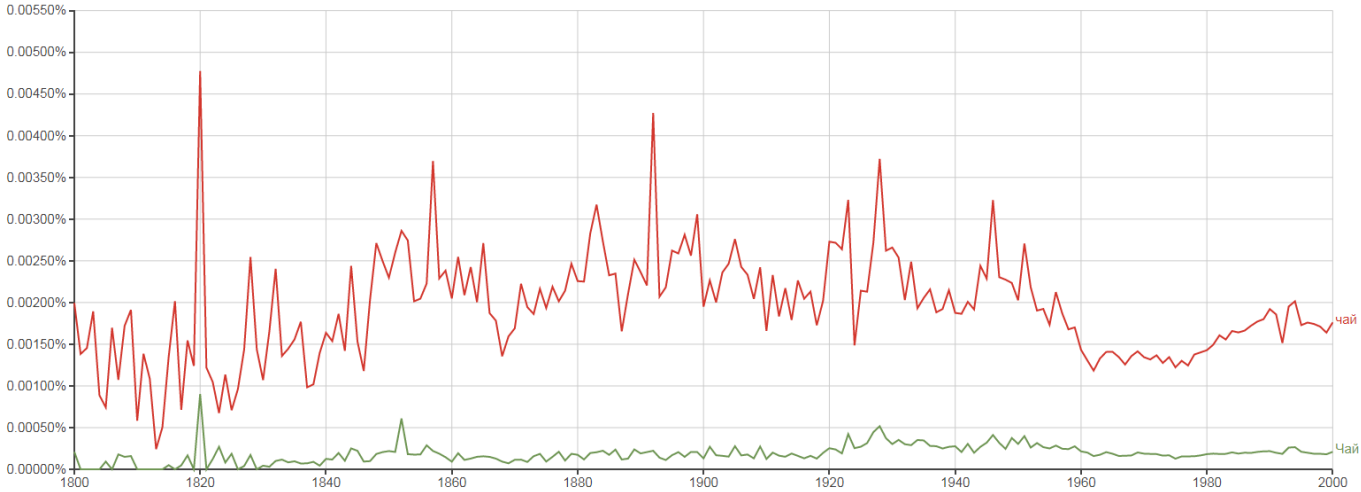
German



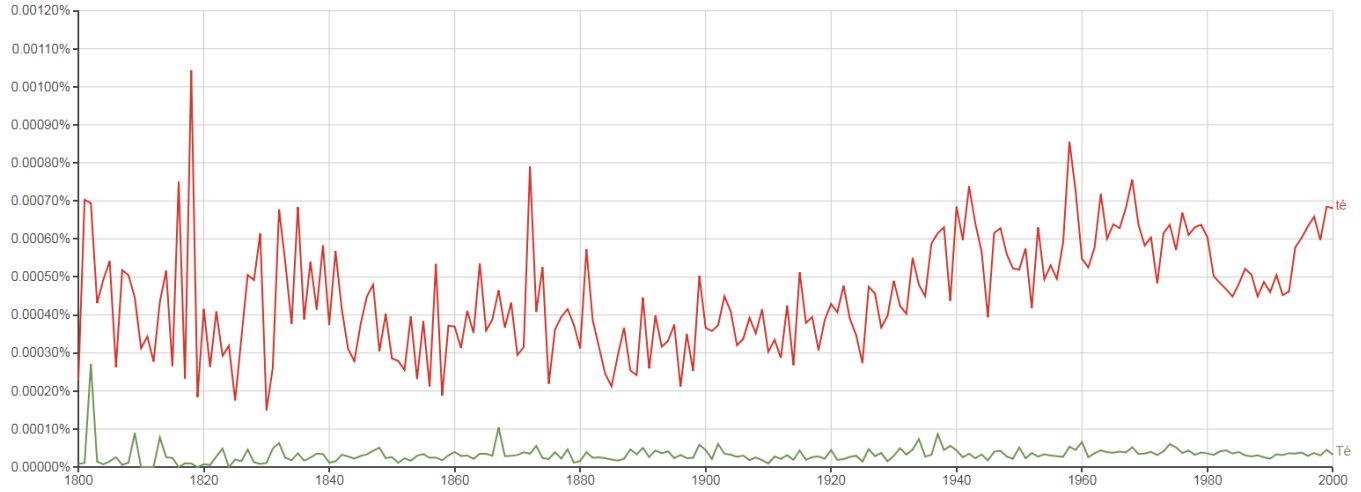
Italian



Russian



Spanish



UK

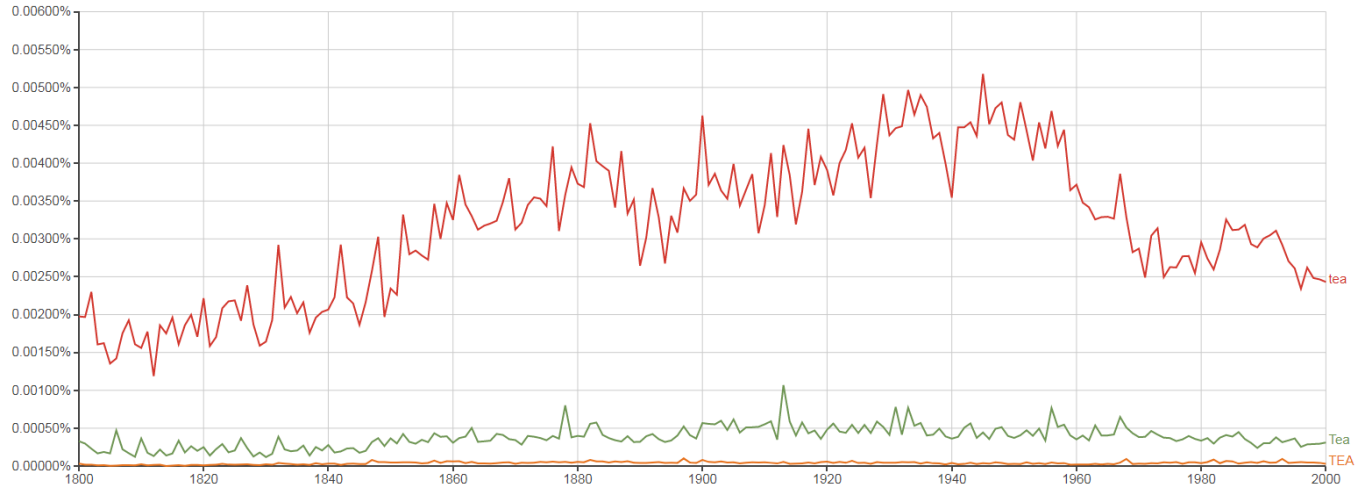
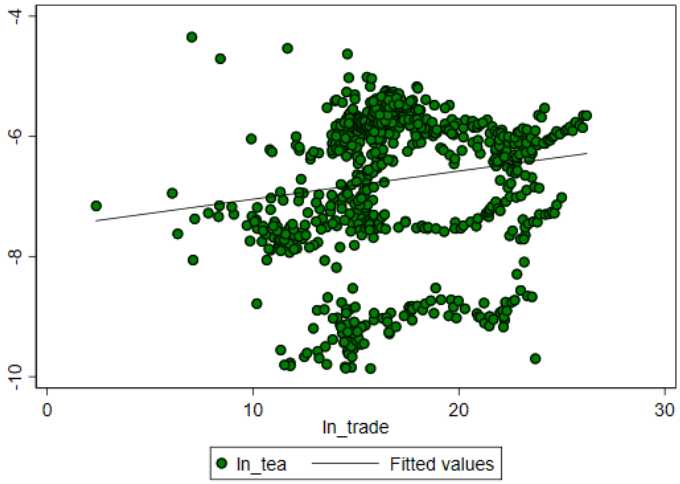
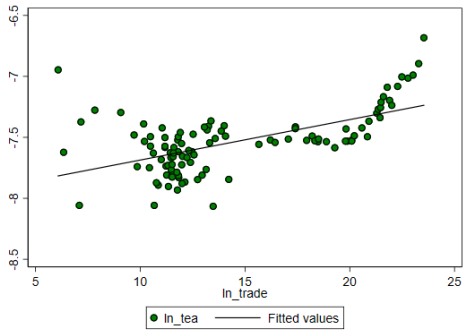


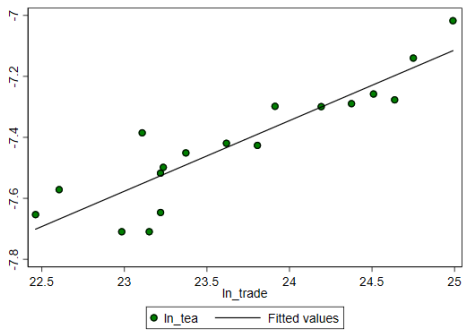
Figure 1: Linguistic frequencies for tea (Google Trends)



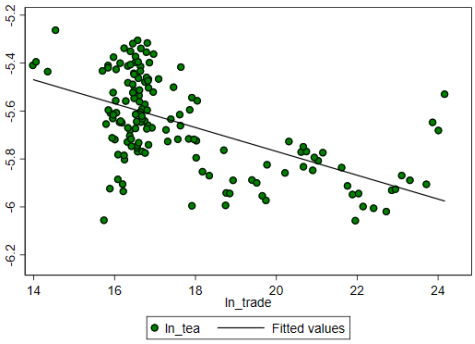
ALL



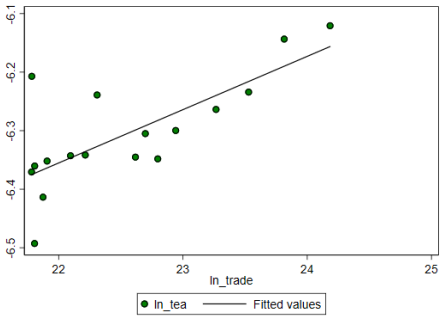
ESP



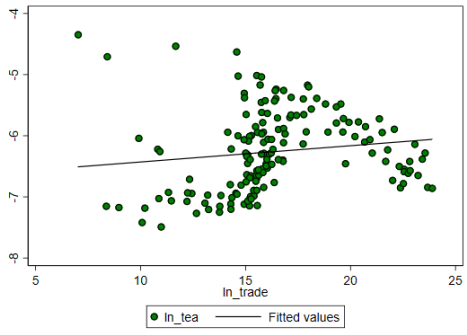
DEU*



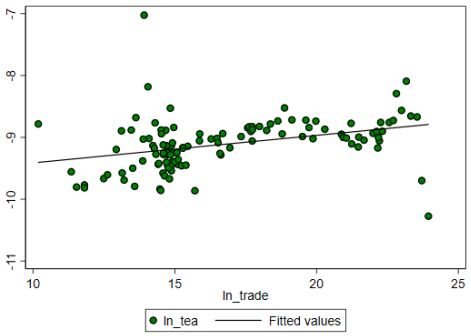
GBR



RUS



FRA



ITA

Figure 2: Correlations between trade flows and the linguistic frequency for tea

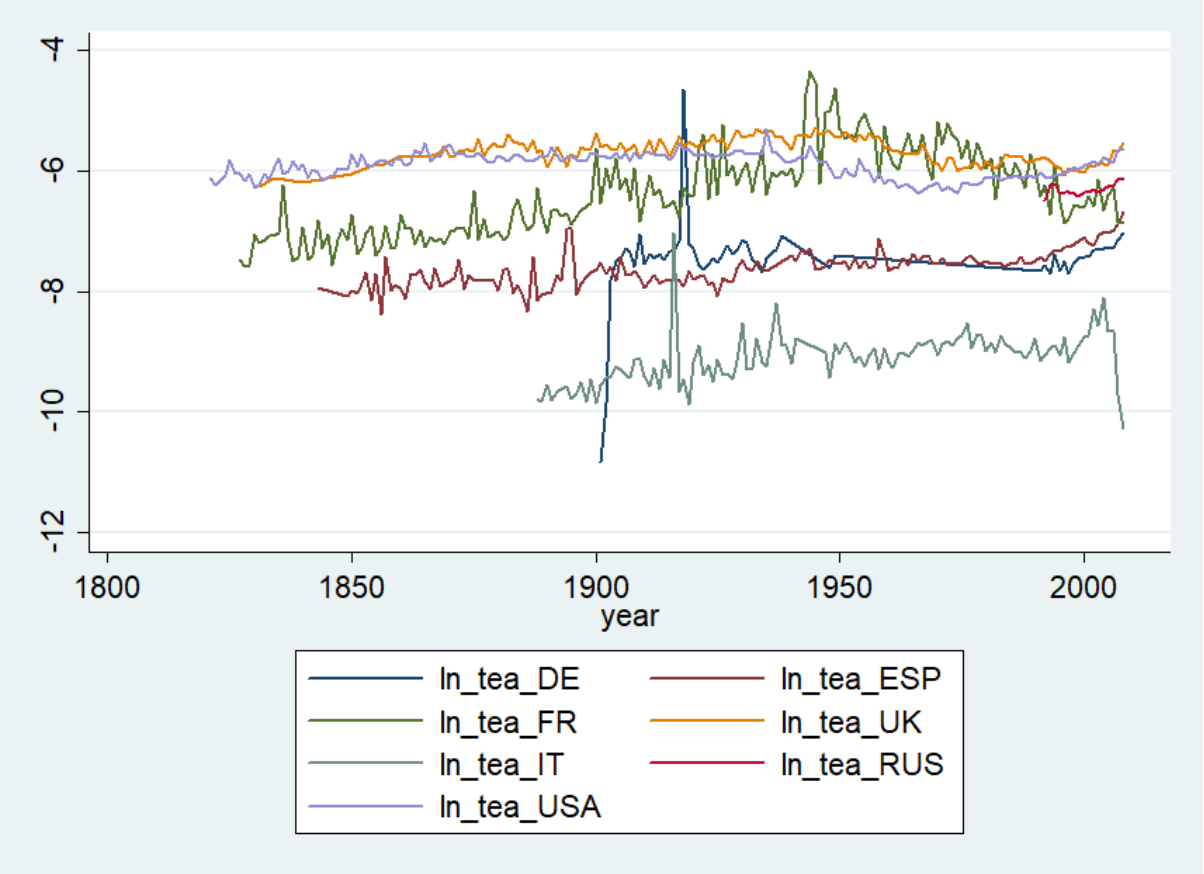


Figure 3: Time series of the natural logarithm of the frequency for 'tea' per country

Table 1: Direct effect of language frequency on trade inflows and outflows (Pooled)

Variables	(1) ln_flow_to_China	(2) ln_flow_from_China
ln_tea	1.429***	0.915***
	(0.155)	(0.146)
ln_GDPi	1.398***	0.066
	(0.195)	(0.176)
ln_GDPj (CHINA)	0.010	1.599***
	(0.278)	(0.265)
ln_distance	2,019.527***	1,827.600***
	(361.782)	(345.562)
ln_dist_sq	-110.395***	-100.135***
	(19.794)	(18.919)
ln_sea_dist	-152.139	-447.900***
	(141.333)	(138.186)
ln_sea_dist_sq	7.687	22.645***
	(7.103)	(6.944)
Contiguity	29.313***	21.987***
	(6.097)	(5.735)
Constant	-8,487.061***	-6,139.858***
	(1,519.913)	(1,436.431)
Year FE	YES	YES
Country FE	YES	YES
Observations	583	604
R-squared	0.930	0.918

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents pooled cross-sectional estimations regarding the effect of language on trade inflows and outflows from China.

Table 2: Direct effect of language frequency on trade inflows and outflows (Panel FE)

Variables	(1) ln_flow_to_China	(2) ln_flow_from_China
ln_tea	0.901***	0.282
	(0.148)	(0.176)
ln_GDPi	0.575***	-0.034
	(0.083)	(0.099)
ln_GDPj (CHINA)	0.769***	1.306***
	(0.098)	(0.116)
ln_distance	-	-
ln_dist_sq	-	-
ln_sea_dist	-162.429	-377.509**
	(138.533)	(169.866)
ln_sea_dist_sq	8.110	18.806**
	(6.953)	(8.527)
Contiguity	-	-
Constant	802.708	1,880.422**
	(689.229)	(845.009)
Observations	583	604
R-squared	0.879	0.789
Number of countries	7	7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents our panel estimation with fixed effects for the effect of language on trade inflows and outflows from China.

Table 3: Direct effect of language on total trade flows (Pooled and Panel)

Variables	(1) ln_trade	(2) ln_trade
ln_tea	0.378**	0.378**
	(0.160)	(0.158)
ln_GDPi	0.183*	0.183**
	(0.095)	(0.088)
ln_GDPj (CHINA)	1.058***	1.058***
	(0.102)	(0.103)
ln_distance	-120.168	-
	(259.792)	
ln_dist_sq	5.239	-
	(14.276)	
ln_sea_dist	-312.828***	-312.828**
	(116.171)	(150.952)
ln_sea_dist_sq	15.578***	15.578**
	(5.834)	(7.577)
Contiguity	-14.385***	-
	(3.715)	
Year FE	YES	NO
Constant	2,220.528**	1,558.721**
	(1,038.858)	(750.942)
Observations	597	597
R-squared	0.845	0.826
Number of countries		7

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the effect of language on overall trade flows to and from China, using pooled OLS with year fixed effects (Specification (1)) and panel estimator with fixed effects (Specification (2)).

Table 4: Direct effect of trade inflows and outflows on language (Pooled)

Variables	(1) ln_tea	(2) ln_tea	(3) ln_tea	(4) ln_tea
ln_flow_to_China	0.074***			
	(0.010)			
ln_flow_from_China		0.039***		
		(0.015)		
ln_GDPi			-0.228***	-0.208***
			(0.036)	(0.049)
ln_GDPj (CHINA)			0.334***	0.301***
			(0.057)	(0.074)
ln_distance			-164.741***	-1,615.980***
			(11.645)	(85.047)
ln_dist_sq			9.217***	87.890***
			(0.646)	(4.642)
ln_sea_dist				19.414
				(34.027)
ln_sea_dist_sq				-0.986
				(1.711)
Contiguity				-26.431***
				(1.587)
Year FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
Constant	-8.753***	-8.243***	725.468***	7,320.146***
	(0.150)	(0.201)	(52.379)	(357.927)
Observations	768	798	661	661
R-squared	0.933	0.923	0.920	0.920

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents pooled cross-sectional estimation with regard to the effect of trade inflows and outflows on language frequency over time.

Table 5: Direct effect of trade inflow and outflow on language (Panel FE)

Variables	(1) ln_tea	(2) ln_tea	(3) ln_tea	(4) ln_tea
ln_flow_to_China	0.026***			
	(0.004)			
ln_flow_from_China		0.013***		
		(0.004)		
ln_GDPi			0.057**	0.035
			(0.022)	(0.023)
ln_GDPj (CHINA)			-0.025	-0.003
			(0.027)	(0.027)
ln_distance			-	-
ln_dist_sq			-	-
ln_sea_dist				-16.237
				(37.098)
ln_sea_dist_sq				0.789
				(1.863)
Contiguity				-
Constant	-7.144***	-6.963***	-7.588***	75.628
	(0.058)	(0.066)	(0.167)	(184.447)
Observations	768	798	661	661
R-squared	0.065	0.014	0.071	0.098
Number of countries	7	7	7	7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the effect of trade inflows, outflows, and their determinants on the frequency of the word tea, using panel estimator with fixed effects.

Table 6: Effect of total trade on language (Pooled and Panel)

Variables	(1) ln_tea	(2) ln_tea
ln_trade	0.016***	0.300***
	(0.004)	(0.027)
Year FE	YES	NO
Constant	-7.000***	-10.276***
	(0.070)	(0.370)
Observations	792	792
R-squared	0.018	0.252
Number of countries	7	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the effect of total trade flow on language, using panel estimator with fixed effects

Table 7: Lagged effect of language on trade inflows and outflows (Panel)

VARIABLES	(1) ln_flow2ch	(2) ln_flowfromCH	(3) ln_trade
ln_tea	0.884***	0.134	0.430
	(0.253)	(0.292)	(0.271)
L.ln_tea	0.194	0.133	-0.022
	(0.236)	(0.274)	(0.253)
ln_gdpi	0.569***	-0.042	0.180**
	(0.084)	(0.098)	(0.089)
ln_gdpj	0.776***	1.308***	1.062***
	(0.098)	(0.115)	(0.104)
o.ln_distance	-	-	-
o.ln_distsq	-	-	-
ln_seadist	-108.287	-259.417	-222.277
	(155.856)	(186.002)	(167.316)
ln_seadistsq	5.388	12.868	11.026
	(7.825)	(9.339)	(8.401)
Constant	534.870	1,293.899	1,109.138
	(775.155)	(925.068)	(832.133)
Observations	574	594	589
R-squared	0.879	0.792	0.825
Number of countries	7	7	7

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the effect of language frequency (of the word 'tea') and lagged language frequency on trade inflows, outflows, and total trade flows with China, using panel estimator with fixed effects

Table 8: Lagged effect of total trade on language (Pooled and Panel)

	(1)	(2)	(3)	(4)
VARIABLES	ln_tea	ln_tea	ln_tea	ln_tea
ln_flow_to_China	0.004			
	(0.019)			
L.ln_flow_to_China	0.021			
	(0.019)			
ln_flow_from_CH		-0.018		
		(0.018)		
L.ln_flow_from_CH		0.030		
		(0.019)		
ln_trade			-0.027	
			(0.020)	
L.ln_trade			0.043**	
			(0.021)	
ln_gdpi				0.730***
				(0.179)
L.ln_gdpi				-0.707***
				(0.181)
ln_gdpj (CHINA)				0.008
				(0.110)
L.ln_gdpj (CHINA)				-0.006
				(0.111)
o.ln_distance				-
o.ln_distsq				-
ln_seadist				-15.612
				(39.625)
ln_seadistsq				0.756
				(1.991)
Constant	-7.118***	-6.925***	-6.985***	72.810
	(0.057)	(0.066)	(0.070)	(196.977)
Observations	739	773	769	633
R-squared	0.065	0.014	0.022	0.120
Number of countries	7	7	7	7

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the effect of inflows, outflows, total trade flows and their lagged values and gravity equation determinants on the frequency of the word tea, using panel estimator with fixed effects

Appendix 1: Descriptive statistics

Variable	Definition	Source	Obs	Mean	Std. Dev.	Min	Max
TEA	frequency of the use of the word tea in China's partner country	Google n-Grams	864	0.0019	0.0016	0.00002	0.0129
flow_to_China	trade inflows from a partner country to China	CEPPI	844	2300000000	8960000000	0	94000000000
flow_from_China	trade outflows from China to the partner country	CEPPI	850	5550000000	26100000000	0	2.84E+11
TRADE	sum of inflows and outflows of trade with China	derived by authors	841	7920000000	34500000000	0	3.78E+11
GDP_o	GDP of partner country	CEPPI	906	4.22E+11	1.25E+12	146000000	1.07E+13
GDP_d_CHN	GDP of China	CEPPI	703	4.72E+11	1.22E+12	779000000	6.29E+12
Dist	road distance between China and partner country in km	CEPPI	900	9885.171	1460.012	5506.708	12298.92
SeaDist	sea distance between China and partner country in km	CEPPI	906	17451.71	4333.848	11280.53	26624.35
Contig	contiguity of the partner country with China	CEPPI	900	0.03	0.16	0	1
TARIFF_o	export tariffs on partner country	CEPPI	645	11.61	10.52	0	57.06
ln_tea	natural logarythm of variable TEA	derived by authors	864	-6.80	1.22	-10.83	-4.35
ln_flow2ch	natural logarythm of variable FLOW_2CHN	derived by authors	810	15.95	4.44	2.14	25.27
ln_flowfromCHN	natural logarythm of variable flow_fromCHN	derived by authors	840	16.56	4.06	2.38	26.37
ln_trade	natural logarythm of variable TRADE	derived by authors	834	17.17	4.01	2.38	26.66

ln_gdpi	natural logarythm of variable GDP_o	derived by authors	906	23.24	3.01	18.80	30.00
ln_gdpj	natural logarythm of variable GDP_d_CHN	derived by authors	703	23.62	2.86	20.47	29.47
ln_distance	natural logarythm of variable Dist	derived by authors	900	9.19	0.15	8.61	9.42
ln_seadist	natural logarythm of variable SeaDist	derived by authors	906	9.74	0.25	9.33	10.19
ln_tarrif	natural logarythm of variable TARIFF_o	derived by authors	644	1.86	1.40	-3.00	4.04
DEU	country dummy - Germny	CEPPI	64	0.07	0.26	0	1
ESP	country dummy - Spain	CEPPI	157	0.17	0.38	0	1
FRA	country dummy - France	CEPPI	188	0.21	0.41	0	1
GBR	country dummy - Great Britain	CEPPI	158	0.17	0.38	0	1
ITA	country dummy - Italy	CEPPI	122	0.13	0.34	0	1
RUS	country dummy - Russia	CEPPI	23	0.03	0.16	0	1
USA	country dummy - United States of America	CEPPI	194	0.21	0.41	0	1

Appendix 2: Tests for fixed or random panel estimator

	ln_flow_to_China	ln_flow_from_China	ln_trade
Breusch and Pagan Lagrangian multiplier test for random effects	Test: Var(u) = 0 chibar2(01) = 1093.71 Prob > chibar2 = 0.0000	Test: Var(u) = 0 chibar2(01) = 216.00 Prob > chibar2 = 0.0000	Test: Var(u) = 0 chibar2(01) = 245.09 Prob > chibar2 = 0.0000
Hausman test Fixed versus random effects	Ho: difference in coefficients not systematic chi2(3) = = (b-B)'[(V_b-V_B)^(-1)](b-B) = 247.57 Prob>chi2 = 0.0000 (V_b-V_B is not positive def.)	Ho: difference in coefficients not systematic chi2(3) = = (b-B)'[(V_b-V_B)^(-1)](b-B) = 79.14 Prob>chi2 = 0.0000 (V_b-V_B is not positive def.)	Ho: difference in coefficients not systematic chi2(3) = = (b-B)'[(V_b-V_B)^(-1)](b-B) = 82.89 Prob>chi2 = 0.0000 (V_b-V_B is not positive def.)

Note: The table presents the test for OLS versus a panel estimator (Breusch and Pagan test) and the test for panel fixed effects versus panel random effects (Hausman tests) for the three types of trade data: inflows, outflows, and total trade flows. The model used for estimation is a reduced form of model (1), where trade flow is explained with the frequency of the word tea and the standard explanatory variables in a gravity model, gdpi, gdpj, and distance.

Appendix 3: Robustness Check 1 – Different Time Cut off Periods

	(1)	(2)	(3)	(4)
	if year > 1900	if year > 1925	if year > 1950	if year > 1975
VARIABLES	ln_flow_to_China	ln_flow_to_China	ln_flow_to_China	ln_flow_to_China
ln_tea	0.959***	1.078***	1.138***	0.139
	(0.197)	(0.200)	(0.204)	(0.090)
ln_gdpi	1.254***	0.223	-0.598**	-0.142
	(0.302)	(0.273)	(0.298)	(0.154)
ln_gdpj	0.047	1.063***	2.303***	1.574***
	(0.298)	(0.262)	(0.278)	(0.121)
ln_distance	1,732.033***	1,579.280***	1,249.229***	109.185
	(464.251)	(399.800)	(391.765)	(2599594.548)
ln_distsq	-95.000***	-87.116***	-69.237***	-6.915
	(25.520)	(21.919)	(21.448)	(142,855.023)
ln_seadist	-371.919*	-555.338***	-591.557***	-585.812
	(209.147)	(162.387)	(150.496)	(1314839.064)
ln_seadistsq	18.904*	28.062***	29.841***	29.759
	(10.530)	(8.176)	(7.577)	(67,063.166)
Contiguity	23.057***	14.908**	7.210	-6.898
	(6.991)	(6.274)	(6.344)	(33,244.529)
Constant	-6,070.392***	-4,414.471***	-2,719.398*	2,444.948
	(1,721.245)	(1,578.773)	(1,596.956)	(5382263.768)
Year FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
Observations	416	344	306	200
R-squared	0.921	0.933	0.932	0.975

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the re-estimation of the effects for OLS with pooled cross-scetion with different cut off intervals. Namely, we have kept the observations only from a different cut off point on-wards – respectively these are cut off points: year 1900 (for column 1), year 1925 (for column 2), year 1950 (for column 3) and year 1975 (for column 4). Similar tables are available for all estimations upon request.

Appendix 4: Robustness Check 2 – Excluding Germany and Russia

	(1)	(2)	(3)	(4)
	if year > 1900	if year > 1925	if year > 1950	if year > 1975
VARIABLES	ln_flow_to_China	ln_flow_to_China	ln_flow_to_China	ln_flow_to_China
ln_tea	1.543***	1.132***	1.189***	0.138
	(0.256)	(0.220)	(0.224)	(0.103)
ln_gdpi	1.154***	0.210	-0.756**	-0.237
	(0.350)	(0.328)	(0.367)	(0.257)
ln_gdpj	0.256	1.106***	2.453***	1.634***
	(0.346)	(0.312)	(0.336)	(0.187)
ln_distance	-191.526	519.353	-236.784	-736.612
	(1,459.466)	(1,182.644)	(1,124.130)	(4262905.178)
ln_distsq	8.407	-29.961	10.978	39.388
	(79.381)	(64.306)	(61.105)	(232,147.990)
ln_seadist	-381.795*	-555.132***	-586.805***	-270.010
	(219.421)	(174.085)	(161.349)	(1037401.358)
ln_seadistsq	19.360*	28.048***	29.598***	13.662
	(11.048)	(8.765)	(8.124)	(52,831.618)
Continuity	ommitted for collinearity	ommitted for collinearity	ommitted for collinearity	ommitted for collinearity
Constant	2,926.930	497.210	4,138.502	4,759.510
	(5,715.637)	(4,676.200)	(4,493.155)	(14476827.886)
Year FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
Observations	364	304	271	165
R-squared	0.913	0.921	0.920	0.971

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents the re-estimation of the results from Appendix 3, but excluding the observations for the countries Germany and Russia. These were omitted because Russia had a very short series of observations and Germany started from a later period than all other countries.