University of Toronto
Department of Economics

Working Paper 691

How Cultures Converge: An Empirical Investigation of Trade and Linguistic Exchange

By Arthur Blouin and Julian Dyer

February 25, 2021

How Cultures Converge: An Empirical Investigation of Trade and Linguistic Exchange

Arthur Blouin
University of Toronto

Julian Dyer
University of Exeter

February 24, 2021*

This paper empirically investigates whether potential gains from trade influence cultural convergence. We develop a model of linguistic convergence where individuals adopt another language to facilitate trade with that group, and diffuse elements of the language throughout their own culture. The model maps to a dataset of linguistic adoption, featuring nearly all words in all languages. We construct a society-pair measure of language adoption that we show is related to welfare gains from agricultural trade. In particular, we show empirically that (1) improved trade-partner quality can cause cultural convergence; (2) adoption is inverse-U shaped in the quantity of trade-partners; (3) economic leverage determines the direction of convergence. We also provide evidence that the language adoption we identify is cultural rather than purely functional by showing that religious and social organization word-types (amongst others) are heavily adopted.

Keywords: Language Evolution; Linguistic Distance; Linguistic Exchange; Loanwords; Trade Incentives.

JEL Codes: O11, O12, O13.

*We are grateful to Francesco Amodio, Gustavo Bobonis, Loren Brandt, Shari Eli, Benjamin Enke, James Fenske, Per Fredriksson, Julian Jamison, Nicholas Li, Rocco Macchiavello, Martina Miotto, Peter Morrow, Sharun Mukand, Naomi Nagy and Jordan Roper for helpful comments. Frederick Gietz, Dina O'Brien, and Matthew Schwartzman provided outstanding research assistance. We also thank seminar and workshop participants at ASREC, CAGE Summer School, Econometric Society Meetings, H2D2, Columbia University, Laval University, University of Toronto, and University of Warwick. We gratefully acknowledge financial support from SSHRC and the Connaught fund.

1. Introduction

Jeffrey Garten, a former U.S. Undersecretary of Commerce for International Trade, once called free trade "a Trojan Horse...that would dominate foreign lifestyles and values" (Garten, 1998). This is consistent with the surprisingly negative public opinion on free trade, especially among those with a strong local identity (Mayda and Rodrik, 2005). Data collected by Silver et al. (2020) highlight that people commonly fear that "globalization [is] breaking down the national community and changing what it means to be part of the nation-state," and that "openness to foreign ideas and customs [is] changing their country's culture." These cultural concerns have contributed to political support for large policy shifts towards economic nationalism, at significant economic cost (Fajgelbaum et al., 2020). As evidence begins to mount that culture and identity can shift quickly (Blouin and Mukand, 2019; Atkin et al., forthcoming), fears that trade will erode culture already impact trade policy (Belluzzi, 1995; Kish, 2001; Maltais, 2016; Grossman and Helpman, 2020; Shayo, 2020). Are these fears warranted? There is actually very little evidence on whether trade influences culture (Bisin and Verdier, 2014), but this paper aims to fill that gap.

We focus on language, and specifically loanwords - words adopted from other languages - as a proxy for cultural convergence.[1] Even beyond the fact that language is a fundamental part of culture, loanword adoption has been heavily interpreted as a direct indicator of culture transmission by linguists and historians for nearly 100 years (Bloomfield, 1933; Scotton and Okeju, 1973; Frankopan, 2016). Notably, societies concerned with cultural erosion are often primarily concerned with language erosion (Schmid et al., 2004), and conversely, linguists consider fears of cultural erosion to be a key factor limiting loanword adoption (Haspelmath and Tadmor, 2009a).

One advantage to our loanwords focus is that cross-cultural language adoption can be accurately and consistently measured for all societies, globally.[2] While linguists have identified about 20,000 loanwords to date, we use a machine learning algorithm to accurately identify the loanword status of nearly every word in nearly every language, a dataset of about 625 trillion word-pairs.[3] Aggregating this word-pair data to the society-pair level - the unit of observation for much of our analysis -

[1]There is a 150 year-old literature in linguistics that aims to distinguish between cognate words, which are inherited from parent languages; loanwords, which are adopted from neighbouring languages; and native words.

[2]The alternative would be to rely on textual sources, which would restrict us mainly to European languages with an extensive written history.

[3]The algorithm that we employ is over 98% accurate.

allows us to construct the intensity of loanword adoption between any two societies. Because of the directional nature of loanwords, we are able to go beyond whether two languages are similar to each other, and examine who adopts language from whom. That is, we are able to separately observe a society's *linguistic borrowing* from another society, and conversely their *linguistic lending* to that society.

We leverage these data to investigate whether loanwords are associated with economic trade. Our empirical work is guided by a model of loanword adoption, based on Lazear (1999). We focus on two actions that lead to societal loanword adoption: some individuals decide to become bilingual to facilitate trade, and then they diffuse words from that language throughout their own society. The bilingualism decision depends on whether an individual's gains from trade are enough to overcome a fixed cost of learning a language. Accordingly, bilingualism is increasing in gains from trade. Diffusion however, is not. Diffusion of foreign words may take place if two bilingual speakers find it useful to use words from a foreign language in conversation. Foreign words would only be helpful if the two speakers were bilingual in the same language, which means that the concentration of bilingualism in a language matters for diffusion. Beyond a certain point, if the distribution of bilingual individuals is spread across too many languages, conversations between those bilingual in the same language are limited, so the diffusion of potential loanwords will not occur.

This framework generates three empirical predictions that we can bring to the data. First, for the predominant foreign language, more gains from trade implies more loanwords usage. Second, loanword adoption by a society has an inverse-U relationship with the quantity of viable trade partners. Finally, for any given trade relationship, borrowing is asymmetric - those who gain more from trade bear the cost of language adoption.

Guided by these predictions, we investigate whether the use of loanwords across societies is associated with economic trade. Drawing on insights from Costinot and Donaldson (2012) and Galor and Özak (2015), we are able to approximate local agricultural gains from trade based on soil characteristic complementarity, if societies prioritize their nutritional requirements.[4] We combine data on potential production of most crops - which we interpret as nutritional endowments - with information on human nutritional requirements. This allows us to estimate welfare in a simple Ricardian model with multiple goods (crops) and a single factor (agricultural land) both under (a) free trade and (b) the counterfactual

[4]Our focus on nutrition helps to isolate exogenous shifters of demand for crops, and sidestep the endogeneity of location-based tastes (Atkin, 2016).

scenario(s) where each society can trade with all but one neighbour (for each neighbour).[5] This trade model accurately predicts contemporary production of all locally traded crops. For each neighbour, we interpret the percentage welfare gained under free trade as *gains from trade*, and also construct *trade influence*, which is gains from trade from the reversed perspective.[6]

With data on both linguistic and economic exchange in hand, we test the model's empirical predictions. We show that a trading partner's quality predicts language exchange. In terms of magnitude, the gains from trade with a typical society's best local agricultural trade partner accounts for about 10% of the regional loanword adoption of a typical society. Even within this relatively narrow focus on local agricultural trade, the evidence confirms that trade incentives do homogenize cultures, which highlights that cultural trade frictions are endogenous just as physical trade costs are (Krugman, 1991). We also show that a society's gains from trade are *only* associated with linguistic borrowing, and *not* with linguistic lending. This helps to highlight that cultural convergence is the result of active economic decisions rather than a passive function of cross-societal contact.

Second, we investigate the implications of partner quantity. Our model suggests a non-linearity in loanwords for the number of viable trading partners a society has. More bilingual speakers means more opportunities for loanword diffusion, but only when they are bilingual in the same language. This intuition is confirmed in our data as well: according to the estimates, linguistic borrowing peaks at 4.75 viable agricultural trading partners. This implies that protectionism - often intended to preserve culture - can sometimes *strengthen* the homogenizing effect of trade on culture.

Finally, we empirically confirm the prediction that language exchange is heavily asymmetric. On average, within a society-pair, the party who linguistically borrows most is the one who gains the most from economic trade. It takes only a 10% difference in gains from trade to induce one society to be the only one to converge in a typical relationship. This may provide some further insight into the implications of cultural protection policies. For example, since language exchange is typically pretty one-sided, a society trading with a new partner would not necessarily imply adoption of that language. This, again, suggests a fairly qualified role for cultural protectionism.

[5] See Feenstra (2004) for a discussion of this classic model, first introduced in Ricardo (1817).

[6] To be more precise. For neighbours i and j, utility under free trade is u_i^{FT}; and under the counterfactual where they can trade with all but one neighbour it is u_i^{FT-j}. Gains from trade is $\frac{u_i^{FT}-u_i^{FT-j}}{u_i^{FT}}$, and conversely *trade influence* is $\frac{u_j^{FT}-u_j^{FT-i}}{u_j^{FT}}$.

The causal argument for the observed association between language exchange and trade is predicated on a key assumption: gains from trade must not influence language other than through actual trade.[7] To demonstrate this, we show that there is no correlation between gains from trade and language exchange for societies that are not viable trade partners. Further, we investigate several alternative mechanisms. One prominent theory of loanwords is the contact hypothesis, which argues that homogenization is an unintended byproduct of interaction, economic or otherwise (Gumperz and Wilson, 1971). The data however, are not consistent with the predictions of this hypothesis. We also investigate colonialism, and find that colonial intensity predicts colonial loanwords, but is orthogonal to gains from trade. Relatedly, we explore conflict, and while we might have expected large gains from trade to correlate heavily with conquest, this is not the case in the data. Similarly, we investigate migration, and find that it is unlikely to explain the patterns in our data. Finally, we control for land diversity, which has been shown to influence ethnic diversity (Michalopoulos, 2012).

One question that remains is whether the language adoption that we identify is, in fact, cultural and goes beyond the bare necessities to facilitate trade. One feature of our dataset is that it is built from word-pair level variation. We can go back to this disaggregated data to explore the types of words being exchanged. We find substantial adoption for each of political, religious, human rights, and social organization word types. We also find evidence of adoption of words for different crops and those relating to economic transactions (like *currency* or *contract*).

Our analysis contributes to a literature on trade and culture within economics, as well as the broader literature on the origins of cultural characteristics.[8] One main contribution is to highlight how linguistic distance is endogenous to trade. Ashraf and Galor (2013a) also explore endogenous diversity with a focus on genetics, while Michalopoulos (2012) considers the inter-generational implications of migration. Meanwhile, Ahlerup and Olsson (2012) theoretically explore the decision to split from a group and Dickens (2019) empirically examines a similar decision.[9] To our knowledge, we are the first to empirically estimate cultural homogenization rather than inter-generational cultural drift.[10] Our results on

[7]A second identifying assumption is that language exchange does not influence the degree of complementarity in soil characteristics among neighbours, which seems plausible.

[8]There is work explaining the origins of cultural differences, see: Nunn and Wantchekon (2011); Alesina et al. (2013); Lowes et al. (2016); Blouin (2020).

[9]Dickens (2019) examines Swadesh lists - words most likely to be vertically transmitted. Changes in these words are likely due to linguistic drift after subgroups breaking away from the main group.

[10]We believe we are among the first to empirically estimate purely horizontal transmission.

partner quantity are also consistent with a literature that has typically found that diversity has quite dire economic consequences.[11] While much of that literature has focused on cooperation (i.e. public goods provision, conflict), our data suggest that beyond a point diversity hinders cultural convergence and cements cultural trade costs even when groups are fully cooperative.

The second literature that we relate to is on trade and culture. One notable line of work in this stream links economic trade with tolerance and peace (Jha, 2013; Jha and Shayo, 2019). There is far more work on the role of culture as a trade friction than on trade's role in shaping culture.[12] The size of this imbalance is somewhat surprising. Since it is well established that culture is an important source of trade frictions, understanding the nature of the endogeneity of these frictions could yield insights. The disparity in focus appears to be due to the lack of data measuring cultural change, as the theoretical literature on the question is more active (Kónya, 2006; Olivier et al., 2008; Gabszewicz et al., 2011). We therefore consider our dataset - which measures global, sub-national, directional cultural adoption - to be an important contribution.

2. Loanwords Background

We measure cultural exchange by studying loanwords. A loanword is a word in one language whose sound and meaning enter the language's lexicon because it was adopted from another language. Loanwords are distinct from cognate words, which are inherited (vertically transmitted) from a parent language. So, two language groups can have similar sounding words that mean the same thing either because they share a parent or because one adopted the word from the other. Linguists typically take considerable effort to first distinguish between loanwords and cognates; and then conditional on identifying a loanword, to identify the direction of transmission.

Detailed case studies of individual languages provide evidence that adoption of loanwords among neighbouring languages is closely linked to "the nature and extent of cultural contacts" (Scotton and Okeju, 1973). They argue that adoption

There is a large literature on vertical versus horizontal transmission (Bisin and Verdier, 2001; Tabellini, 2008; Algan and Cahuc, 2010; Algan et al., 2013).

[11] See: Alesina et al. (2003b); Esteban et al. (2012); Spolaore and Wacziarg (2015); Michalopoulos and Papaioannou (2016); Spolaore and Wacziarg; Desmet and Wacziarg (2018); Michalopoulos et al. (2019)

[12] Culture as a transaction cost is both theoretically (Dasgupta and Serageldin, 1999; Glaeser et al., 2002; Durlauf and Fafchamps, 2005; Putnam, 2007; Guiso et al., 2008) and empirically (Hall and Jones, 1999; Rauch and Trindade, 2002; Giuliano et al., 2014; Melitz, 2008; Guiso et al., 2009; Felbermayr and Toubal, 2010; Gokmen, 2017) well established.

is often heavily influenced by the socio-cultural context of a particular group in relation to their neighbours. Even adoption into the core vocabulary of a language can be prevalent with enough contact with another language group. The prevalence of loanwords in a language should therefore be thought of as the result of a socio-cultural process involving the interactions of individual speakers of languages. Accordingly, more loanwords can be viewed as a proxy for reduced cultural distance.[13]

Socio-linguistics has focused heavily the role of contact in reducing linguistic distance. Gumperz and Wilson (1971) provide evidence that contact and geographic proximity usually comes along with at least some adoption and diffusion. However, this focus on proximity and contact has generally left the literature unable to explain some prominent cases.[14] To resolve these puzzles, the focus has largely been on language characteristics and how they influence the within-language diffusion process. Additionally, linguists have focused on class and prestige as factors influencing diffusion (Labov, 1964; Labov and Harris, 1994), but factors like age (Sankoff and Blondeau, 2007), gender (Cameron and Kulick, 2003), ethnicity (Cukor-Avila and Bailey, 2001), and social structure (Paolillo, 2001) have also been considered.

Linguists have devoted significantly more effort to understanding the ancestry of languages and to identifying the age of branches in linguistic family trees (Vansina, 1990). This task requires *excluding* loanwords in order to focus on non-adopted words that are indicative of parent languages and the timing of splits. Towards this end, linguists have identified lists of core meanings that are fundamental to human languages that can be thought of as necessary. These words are unlikely to have been adopted, since all languages likely had to develop or inherit their own word for these meanings.[15] These Swadesh lists (named for Morris

[13]Having noted that, research on loanwords has traditionally been more descriptive, and has typically been towards the aim of a 'complete' understanding of a language. However, recently linguists have begun to go beyond a purely descriptive treatment of loanwords, and have begun asking questions such as 'why are words for body parts rarely adopted but words for objects are?' This turns out not to be as straightforward as one might imagine. For instance, the English word *window* was adopted from Old Norse even though English had previously used the word *eagpyrel* in precisely the same manner (Haspelmath and Tadmor, 2009a).

[14]One puzzle is why Australian languages have failed to adopt and diffuse any "creole" features at all (Heath, 1984). On the other end of the spectrum, the literature has focussed intensely on Japanese-English adoption as "[s]imply put, there is no significant cultural contact between large groups of Japanese and English speakers" (Hoffer, 2002).

[15]These are meanings such as 'man', 'woman', 'sun', 'night', 'eye', 'water', 'fire'. These meanings are essential and would almost certainly exist in any useable language, and are therefore less likely to be adopted. Meanings outside these core concepts (such as for ideas and manufactured objects) are not necessarily an original part of all languages and are more likely to be adopted from another language.

Swadesh) have become the foundation for many measures of linguistic distance used to measure ancestral distance among linguistic groups, like the Automated Similarity Judgment Program (ASJP) (Swadesh, 1950; Wichmann et al., 2016). For the inverse task of identifying loanwords, there is no such list of concepts that can be applied universally across languages since the lending and adoption of words is so heavily influenced by power, economics and cultural openness (Haspelmath and Tadmor, 2009a). These factors – often an inconvenience to linguists with respect to understanding the evolution of languages – may be of importance to economists, and are a focus of this paper.

Historians have long used the existence of loanwords as evidence of exactly these factors. Furthermore, loanwords have been heavily interpreted as indicators of historical cultural transmission. This is often linked to economic and political power, for instance:

> *"Buddhism made sizeable inroads along the principal trading arteries to the west too [...] The rash of Buddhist loan words in Parthian also bears witness to the intensification of the exchange of ideas in this period"* (Frankopan, 2016, p. 32)

We are amenable to an interpretation of loanwords as a broad proxy for the exchange of ideas, and we do find evidence suggestive of this interpretation. However, it is both unnecessary for our hypothesis, and we are sympathetic to the conceptual issues that may arise from conflating cultural and intellectual exchange more broadly. Our focus is to speak to the determinants of cultural convergence between groups within a region, as proxied by loanword adoption. This is consistent with the interpretation of loanwords by both linguists and historians.

3. Conceptual Framework and Predictions

In this section we summarize a model of strategic linguistic adoption, and state three predictions that we test empirically.[16] The model is presented more formally in appendix section A. Our framework is similar to that of Lazear (1999), adapted here to our setting of cross-cultural trade.[17] We aim the analysis towards predictions on three related issues. First, do gains from trade with a potential trade

[16]Our aim here is a simple intuitive exposition that describes our hypothesis in a way that maps to our data and guides empirical decisions, rather than a complete conceptual description of the issue.

[17]Lazear (1999) on the other hand, models repeated matching with individuals in the same society.

partner affect cultural convergence? Secondly, how does increasing the quantity of trade partners impact convergence? Finally, if convergence does take place, who converges towards whom?

3.A. Summary

Towards the aim of guiding and focussing the empirical analysis, it is useful to distinguish individual-level decisions to invest in a second language (we call this bilingualism) and the process by which these individual decisions lead to foreign words being incorporated into a language (diffusion).

We consider an individual who speaks a native language, but has the option of becoming bilingual. Additional profits to bilingual speakers are generated through less costly trade with societies who natively speak the second language. However, there is a fixed cost associated with becoming bilingual. So, the individual will become bilingual in the language that provides them the greatest gains from trade if those gains from trade are greater than the cost of learning. For simplicity we assume that people can become bilingual in only one language.[18]

Diffusion of a foreign word may take place if the participants of a conversation find it useful to use that word. For a word from a foreign language be used in a conversation, that foreign language must be both known by the using party and understood by the receiving party. This implies the likelihood of diffusion of a word is less dependant on bilingualism generally than it is on the intensity of bilingualism in a particular language. If there is only one foreign language, total loanword diffusion is increasing in bilingualism, which is increasing in gains from trade. However, with multiple foreign languages this is ambiguous. Beyond a certain point, if the distribution of bilingual individuals is spread across too many languages, conversations between those bilingual in the same language are limited, so the diffusion of possible loanwords will not occur. For instance, if everyone spoke a different second language there would be no diffusion and therefore no loanwords. This is true even if everyone in the society is bilingual.

Diffusion can in turn influence bilingualism. For a given language, the more diffusion that has already occurred as the result of widespread bilingualism, the smaller is the benefit of adoption. This is because with substantial baseline linguistic overlap even unilingual individuals can operate in a limited sense in the foreign language. In other words, language adoption explicitly reduces trade costs for

[18]The results would still go through as long as we had diminishing returns to new languages. But it simplifies things to only deal with one foreign language per person.

the adopter, but there may also be externalities for unilingual speakers if cultural convergence takes place at the societal level.[19]

3.B. Partner Quality

The tension in the model comes from the idea that when individuals become bilingual in a particular language there are two competing forces. On one hand, if those individuals would have otherwise not been bilingual, then adoption necessarily increases diffusion. However, some individuals may adopt a new language at the expense of adopting a different language, so individual adoption decisions impact both the number of bilingual individuals, and the distribution of bilingual individuals across languages. In particular, if the adoption decision reduces the intensity of bilingualism in another language, this can reduce diffusion. This occurs if number of speakers of *the same* second language is reduced, even if the number of speakers of *any* second language is increased. Only the former matters for diffusion.

This trade-off implies that while greater gains from trade unambiguously increases bilingualism, it only unambiguously increases total *loanwords* if we are considering the foreign language with the most speakers, or the language with the largest gains from trade. Formally,

Prediction 1 (Quality Matters). *Total loanword diffusion is increasing with economic incentives for interaction with a society's best neighbour.*

So, we might expect trade to homogenize cultures, but not ubiquitously. The context matters. There may be convergence, for instance, if a society is intensifying trade (say through signing a free trade deal) with one dominant trading partner. This seems, at least anecdotally, to reflect reality.[20]

[19]These externalities need not be exclusively positive. For instance, consider someone who dislikes another group. Their dislike of members of the other group means they would never trade with that group, so they do not earn the benefits of easier trade, and they may perceive some cultural loss that is valuable to them. However, for our purposes we do not need adoption to be Pareto improving, only that it weakly increases trade for all.

[20]75% of Canadian trade is the the United States, and Canada is famously concerned with trade and cultural protection. Canada's Prime Minister made waves by walking out of Trans-Pacific Partnership talks in 2017 over concerns about cultural protections. But the concern is longstanding. *The New York Times* reported in 1998 that Canada hosted 19 nations, excluding America, to discuss "what they see as the gravest threats to their collective cultures – free trade and the United States" (Depalma, 1998).

3.C. Trade-partner quantity

The other implication of the tension in the model is a non-linearity in partner quantity. The most straightforward way to show this is to assume all of the society's viable trade partners (i.e. when there exist positive gains from trade) have the same quality and therefore the same number of bilingual speakers. We have already highlighted the possibility of crowding-out popular languages. The key thing to note is that crowd-out is necessarily low when there are very few bilinguals, and is necessarily high when there are many.[21] This intuition leads to our second prediction:

Prediction 2 (Quantity Matters)**.** *Total adoption is inverse-U shaped in the number of viable trading partners, while total lending is non-decreasing.*

So, a country that is faced with protectionist pressures due to perceived cultural loss from trade can actually *accelerate* convergence by reducing trade. The result may also have implications for the diversity literature, which typically finds that diversity impedes economic development because it makes cooperation more difficult. Here, having a diverse set of natural trading partners reduces cultural convergence, potentially increasing the cultural costs of trade with each of them - even if everyone is fully cooperative.

3.D. Who converges towards whom?

One question that remains: since only one common language is required to facilitate trade, who bears the fixed cost of becoming bilingual? When a trading relationship is viable, becoming bilingual is a strategic decision that we can model as a war of attrition game. Under standard assumptions for this type of game, the intuitive result is that the party who gains the most from trade (and hence loses the most whenever trade is not facilitated) will be the more likely to incur the cost of becoming bilingual.[22]

Prediction 3 (Asymmetric Incentives Matter)**.** *For a pair of neighbours, the group that gains the most from trade will be the society that bears the cost of acquiring a common language.*

[21] This is simply mechanical. For example, consider a society of 10 people, with 5 new adopters. This cannot crowd-out any other language if we start with no bilingual speakers, but it must crowd out at least 4 if we start with 9 bilinguals.

[22] See Appendix A.2.3 for the full proof of this result, following the solution in Levin (2004) of the two-player case of the Generalized War of Attrition in Bulow and Klemperer (1999) where there is a small chance that a player will, irrationally, never make the adoption investment despite the potential gains from trade.

Finally, the most common alternate framework of cultural convergence in the broader social science literature is the *contact hypothesis*, where cultural adoption follows as an unintended byproduct of interaction. A key distinction between this framework and the one that we outline above is that under the contact hypothesis, bilingualism is only a function of the number of viable trade relationships. This leads to mutual convergence where the only asymmetries result from differences in population size. These predictions are outlined formally in appendix section A.3.

4. Data

4.A. Language Data

i) PanLex In order to construct data on loanwords and linguistic exchange, we need wordlists, or *lexicons*, from as many languages as possible. For this, we draw on the PanLex database, which takes thousands of translation dictionaries converted to a single common structure, covering over 25,000,000 words.[23] The dataset is as close as we believe is possible to representing all known words in all known languages. The coverage of this dataset goes far beyond the coverage possible with sources based on textual and archival resources, which are restricted to languages with a significant body of written history. This breadth of coverage is a further advantage of the loanwords approach.

ii) World Loanword Database We combine PanLex lexicons with information on classified loanwords from the World Loanword Database (WoLD). WoLD is a scientific publication by the Max Planck Institute for Evolutionary Anthropology, and includes 41 recipient languages and 369 donor languages. Figure 1 presents a map of the spatial distribution of each type of language in WoLD. It is the first aggregated dataset of rigorously-identified loanwords under a consistent set of criteria, providing "...vocabularies (mini-dictionaries of about 1000-2000 entries) of 41 languages from around the world, with comprehensive information about the loanword status of each word" (Haspelmath and Tadmor, 2009b). The data compiled into WoLD is the result of a long literature on loanwords by linguists.[24]

[23]PanLex is a non-profit organization with a mandate to build the largest possible lexical translation database with the aim of improving resources available to under-served languages: see https://panlex.org. The database is constantly being updated, in this paper we use the SQL database posted on September 1, 2019

[24]The WoLD data for Swahili, for example, is based on thirty-three academic publications by twenty-seven separate authors, published between 1861 and 2001.

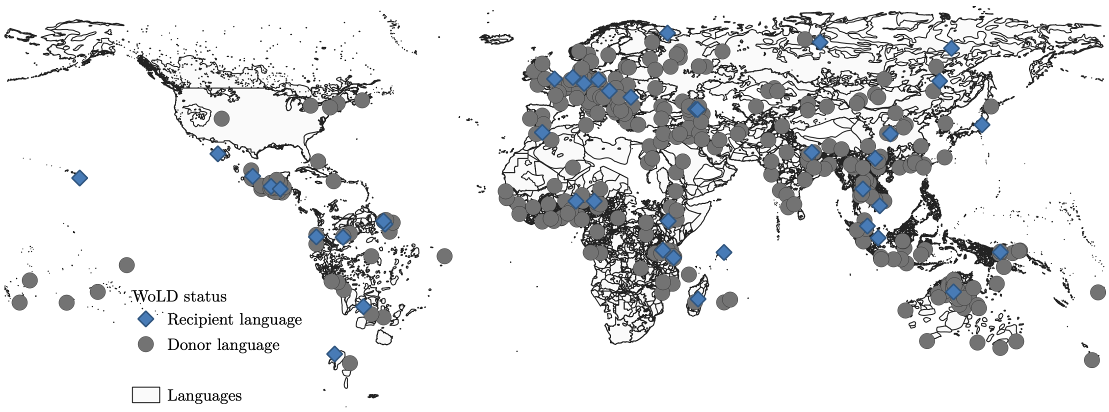
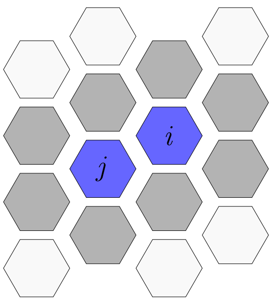
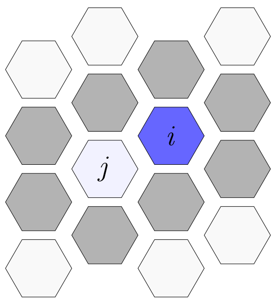


Figure 1: Map of WoLD language groups

Note: This map shows each of the borrower and lender languages in the WoLD dataset. The grey dots represent lending languages while the blue diamonds represent borrower languages. In total there are 395 languages mapped, 41 of which are borrowers and 369 are lenders (this does not add to 395 because 15 languages are both lenders and borrowers).
Source: Author constructed using data from WoLD: https://wold.clld.org/language. Last Accessed December 22, 2020 2:00pm EST.



(a) Full Trade: U_i^{FT}



(b) Counterfactual without j: U_i^{FT-j}

Figure 2: Neighbourhoods Used for Constructing Gains from Trade

Note: This figure illustrates the counterfactual neighbourhoods used for our structural estimates of gains from trade at the language-pair level. A dark shaded polygon indicates a society that is included in the given counterfactual neighbourhood. In panel a) we show the neighbourhood used for our full trade counterfactual between group i and j, made up of the union of immediate neighbours of i and j. In panel b) we show the counterfactual neighbourhood where j is dropped from the neighbourhood that i can trade with.

4.B. Potential Gains from Agricultural Trade

i) Neighbours and Location Data We use the digitized Ethnologue map of ethnolinguistic societies (Lewis, 2009) to define societies and their local neighbours. We construct neighbourhoods that set the scope of possible historical agricultural trade in our models. We consider trade in a neighbourhood constructed as the union of the immediate neighbours of the two societies. See Figure 2 for a graphical representation of the neighbourhoods of interest.

ii) Potential Agricultural Production Data on agricultural productivity comes from the Global Agro-Ecological Zones (GAEZ) dataset (IIASA/FAO, 2012), which includes measures of potential production for 49 crops at the 5 arc-minute grid-cell level for the entire world.[25] We combine this crop productivity data with the Ethnologue to construct society-level averages of long-run potential production of each crop. After matching the Ethnologue, PanLex and GAEZ, and dropping societies without any neighbours that have non-missing observations in all three datasets, we are left with 2,606 societies and 9,436 society-pairs (Table 1).

Table 1: Descriptive Statistics

	Mean (1)	Variance (2)	Min (3)	Max (4)	N (5)
Panel A: Society Pair Level					
Language Adoption	0.26	1.91	0	100	9,436
Gains from trade (percentage change)	0.067	1.24	-2.83	3.34	9,436
Gains from trade (percentile rank)	0.500	0.289	0	1	9,436
Trade Utility	2.61	1.72	0.003	14.27	9,436
Population (1,000)	9,415	72,625	0	871,558	9,436
Area Share of Neighbourhood	0.097	0.137	0	0.999	9,436
Land Diversity	29,518	39,920	0	343,933	9,436
Distance to Neighbour (km)	236.3	459.3	0	5,628	9,436
Panel B: Society Level					
Language Adoption (total)	20.01	16.9	0	100	2,606
Language Adoption (best neighbour)	0.79	3.54	0	100	2,606
Language Influence (best neighbour)	0.63	3.19	0	100	2,606
Gains from trade (best neighbour, pct change)	0.96	1.45	-2.83	3.34	2,606
Trade Influence (best neighbour, pct change)	0.78	1.31	-2.83	3.34	2,606
Gains from trade (best neighbour, pctl rank)	0.73	0.25	0	1	2,606
Trade Influence (best neighbour, pctl rank)	0.69	0.24	0	1	2,606
Trade Utility	2.50	1.70	0.003	14.27	2,606
Population (1,000)	1,389	18,518	0	871,558	2,606
Area Share of Neighbourhood	0.111	0.145	0	0.991	2,606
Land Diversity	24,628	37,913	0	343,933	2,606
Distance to Neighbour (km)	173.4	325.7	5.54	5,085	2,606
Share of viable trading relationships	0.61	0.34	0	1	2,606

Note: The table shows descriptive statistics for the main variables used throughout the empirical analysis. We have word-level data for 9,436 society-pairs. The population data comes directly from the Ethnologue. Distance to neighbour is author constructed based on the Ethnologue centroids. The utility data all comes from a trade model, which is described in section 5.B.

[25]To avoid concerns regarding endogenous irrigation or other agricultural inputs, we use the potential yields for low-input, rain-fed agriculture. This is similar to the methodology used for generating the measures of crop productivity in Galor and Özak (2016).

iii) Nutritional Content and Requirements We use data on the nutritional content of crops, which comes from FAO databases (FAO, 2017a,b) and is matched to the crops in the GAEZ data.[26] To measure the required essential nutrients to sustain the average adult human,[27] we use the Dietary Reference Intakes (DRI) tables produced by the Institute of Medicine, National Academy of Sciences (Institute of Medicine, 2006).

5. Data Processing

All of our analysis is done either directly at the society-pair level, or on society-level data that is aggregated in different ways from society-pair data. We would therefore like to bring together the data sources above to construct a society-pair level dataset that includes both gains from trade and linguistic exchange. This section describes how we use PanLex and WoLD to construct linguistic borrowing and lending at the society-pair level, as well as how we use the GAEZ, DRI, and Ethnologue data to construct gains from trade and trade influence, also at the society-pair level.

5.A. Constructing Society-Pair Level Linguistic Exchange

As discussed in Section 4.A, WoLD is incomplete. While it is quite a large dataset, it covers only a small fraction of PanLex. Ideally, we would like to understand, for every word in every language, whether it is a loanword and where in the world it was adopted from.

To do this we train a machine learning prediction algorithm, which is the only feasible way to accomplish this at the scale required. From PanLex we started by creating a word-pair level database. PanLex includes 25,000,000 words which results in $6.25 \cdot 10^{14}$ (625 trillion) word-pairs.[28,29] From WoLD we had a good un-

[26]Specifically, these data cover twenty-three micronutrients for forty-one of the forty-nine crops included in our agricultural productivity data.

[27]There are sixteen nutrients in our crop content data that are also identified as essential nutrients by feeding experiments in Chipponi et al. (1982), where "[t]he dietary *essentiality* of an organic compound signifies that it serves an indispensable physiological function, but cannot be synthesized endogenously."

[28]Running a machine learning algorithm for a dataset of this size requires considerable computational power. To implement this we relied heavily on SciNet, the largest supercomputer in Canada. The Niagara supercomputer at SciNet is owned by the University of Toronto, and includes a homogeneous cluster of 61,200 cores. Of this we were allocated 13.5 core-years, and our algorithm ran for approximately 43,760 core-hours on every candidate word-pair in PanLex. This took approximately one week using 300 cores. For a rough comparison, this would have taken about 1.25 years on a standard quad-core laptop.

[29]There were some important decisions to make in order to manage computational resources,

derstanding - for a subset of those word-pairs - of whether one word was borrowed from the other, and the direction of transfer. We used this subset of word-pairs whose loanword status was known as a training set. Based on this training set, we then estimated - for all word-pairs with unknown loanword status - whether one word originated from the other.

To do this, we first needed to generate the features of word-pairs from which our classifier could generate predictions.[30] For a potential word pair we generated features that fall into three categories. First, we generated measures that indicate the linguistic similarity of a potentially borrowed word to its own language, since a word that is an outlier relative to its own language may be more likely to have been borrowed. Second, we generated the same own-language similarity measures for the potential source word (i.e the original word that was borrowed as a loanword into another language) since a word that is an outlier in its own language may be less likely to be the source of a transfer. Next, we generated features to measure the phonetic and orthographic similarity of the word-pair, since more similar words may be more likely to have been part of a transfer. Finally, we include a measure of the distance between the two languages in a language family tree, to allow our classifier to take this into account when setting thresholds.

Importantly, we relied entirely on language tree distance, orthographic and phonetic features to make loanwords predictions. Our classifier does not observe variables that are directly indicative of the identities of the languages themselves (such as language family, lexicon size, population, or region, etc.). This means our algorithm is classifying on the characteristics of a word-pair, and not overfitting to simple and potentially problematic rules such as 'Nilo-Saharan languages adopt a lot,' or 'Smaller groups tend to adopt from bigger groups.'

Our second challenge was that the training data are heavily unbalanced. The number of loanword pairs are dwarfed by the number of non-loanwords pairs. This is a potential problem, because we could estimate that there has never, in any language ever been a loanword, and achieve very high accuracy, which is clearly not what we want. We used a combination of two methods to deal with this issue. The simplest is to under-sample the heavily-represented group, the second is synthetic minority oversampling.[31] This provided us with training sets

even though we had access to the supercomputer. For details on the set-up and decisions relating to navigating our computational resources, please see Appendix B.

[30]These features are listed and explained in detail in Appendix B, including a description of how orthographic and phonetic measures were implemented.

[31]'Synthetic' examples of the under-represented type of observation were resampled with replacement. These synthetic examples were constructed as a convex combination of nearest neighbours of the same type within feature space. See Chawla et al. (2002) for a discussion of

Table 2: Evaluating Classifier Performance

Classifier Type	Classifier Performance Measures (on Test-Set)				
	Accuracy Score	F1 Score	Balanced Accuracy Score	Precision Score	Recall Score
	(1)	(2)	(3)	(4)	(5)
Panel A: First-Stage:					
Random Forest	0.9835	0.8386	0.9215	0.8251	0.8526
Extremely Random Forest	0.9751	0.7580	0.8794	0.7437	0.7729
Voting Classifier	0.9836	0.8386	0.9187	0.8308	0.8466
Panel B: Second-Stage:					
Random Forest	0.9125	0.8979	0.9108	0.8968	0.8990
Extremely Random Forest	0.9067	0.8872	0.9005	0.9186	0.8579
Voting Classifier	0.9088	0.8922	0.9055	0.9019	0.8828

Note: Each column presents accuracy scores using a different metric, each weights false-positives relative to false-negatives differently. We show each for both our first and second stage classifiers. All performance measures are performed on the test-set, none of the training-set data is not included in this table. Accuracy is simply the share of word-pairs classified as a loanword, are actually loanwords, with the correct direction of borrowing. The precision score is the share of predicted positives that are true positives. The recall score is the share of actual positives that are predicted. Balanced accuracy is the mean of the true positive rate and the true negative rate. The F1-score is the harmonic mean of precision and recall.

for Random Forest classifiers, as well as an Extremely Randomized Forest, which is conceptually similar but further decreases overfitting.[32]

From these three classifiers we built an ensemble Voting Classifier that is approximately 98% accurate (table 2, panel A and figure 3(a)). This high accuracy score could be misleading if the high overall accuracy comes at a cost of very low accuracy in particular categories. For this reason we also report a number of different diagnostic measures that balance different types of errors. For instance, *precision* is the share of predicted positives that are true positives, while *recall* is the share of actual positives that are predicted. In the scenario described above, if the algorithm drastically under-predicted true positives to achieve high overall accuracy, the recall score would be nearly 0, but the precision could be high. The F1-score and balanced accuracy scores both account for this issue.[33] Our classifier still does well on these additional checks, with an 84% recall score (panel A column 5 and figure 3(b)), and an F1-score of 0.84 (panel A column 2).

In fact figure 3(b) goes a bit further, by showing exactly where our accuracy

the theory of SMOTE over-sampling and see Lemaitre et al. (2017) for the details of the exact implementation used in this paper.

[32]As discussed in Mullainathan and Spiess (2017) and Varian (2014), these ensemble classifiers improve out-of-sample fit by ensuring that the learning algorithm does not over-fit to the training set. To choose the optimal hyperparameters for these classifiers, we used a grid search method over the number of features available at each split of the decision tree, the maximum depth of the decision tree, and the minimum number of observations in each final leaf, and select the parameters that performed best on different folds of the training set.

[33]Balanced accuracy is the mean of the true positive rate and the true negative rate. The F1-score is the harmonic mean of precision and recall.

is coming from on the test-set. The blue line represents our accuracy on word-pairs that are actually loanword-pairs. This is essentially the recall score - on actual loanwords, we are correctly identifying them as loanwords about 84% of the time. Next going bottom to top (on both the figure and legend), we examine the categories of word-pairs that are actually a loanword pair, but the direction of transfer is flipped. In these cases we correctly identify the pair as not being a loanword-pair almost 95% of the time. For words that are not loanwords at all - for instance words that we know are not borrowed at all - we correctly categorize words as non-loanwords over 98% of the time. Finally, we essentially never mis-categorize an actual loanword as being borrowed from the wrong potential source word in another language.

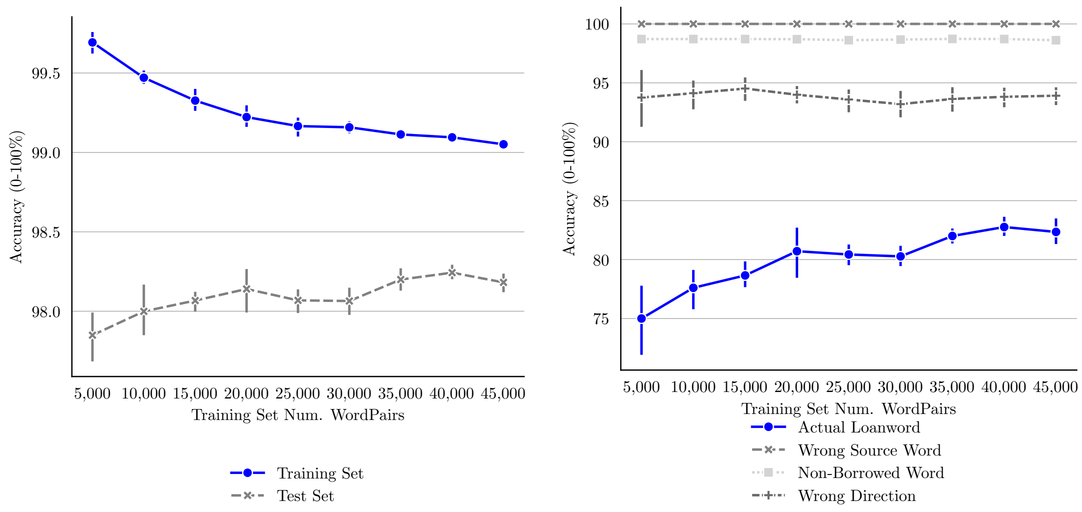
To further ensure that the word-pairs we identify as loanword adoption are not false positives, we trained a second-stage classifier to further filter the pairs that the Voting Classifier (described above) identified as loanword pairs. This second-stage classifier is approximately 91% accurate on a test set of word-pairs identified as being loanwords by the first stage classifier with an improved F1-score of 0.89 (as shown in figure E4 and in table 2 panel B). Machine learning typically requires a lot of data to be effective. To show that our sample size is adequate, we bootstrap the training set at different sizes. Observe that the accuracy of the classifier is no longer increasing as we reach the size of training set we use in our main results (figure 3).

We then applied these classifiers to the full set of possible word-pairs in the PanLex data, exactly as when we constructed the training set. We took these predicted word-pairs, and where two source words (here, source refers to the word in another language which was borrowed as a loanword) were identified for the same loanword, we kept the source word with the highest probability from the second stage classifier.[34] We report descriptive statistics for language exchange measures generated by this methodology in table 1.[35]

i) Construction of Dependant Variables of Interest We then aggregated the word-pair level results into a society-pair level dataset. This results in two sets of variables that we use throughout the analysis. First, at the society-pair level, we

[34]We also drop loanwords where the source word was itself identified as a loanword, so our final measure of language exchange only includes unambiguously identified loanword pairs.

[35]We also show, in the left panel of Figure 4, how this measure is distributed spatially for language groups in Africa, and in figure E3 we present a histogram of language adoption.



(a) Overall Scores

(b) Test-Set Score by Word-pair Type

Figure 3: Accuracy of Voting Classifier

Note: The figure shows the accuracy of the machine learning algorithm by training set size. On the y-axis we show the share of word-pairs classified correctly by the algorithm. We contemplated adding observations to the training set, but the graphs suggest that about the past 10,000 word-pairs have not made meaningful improvements in accuracy. Furthermore, accuracy rates are quite high.

define a measure of linguistic borrowing as follows:

$$\mathcal{L}_{ij} = \frac{\mathbf{card}(Loanword_{ij})}{\mathbf{card}(Word_i)} \tag{1}$$

We define $Word_i$ as the set of words in the language of society i, and so $\mathbf{card}(Word_i)$ is the cardinality of the set of words in the language of i. Similarly $Loanword_{ij}$ is the set of loanwords in the language of society i originating from j, and $\mathbf{card}(Loanword_{ij})$ is the cardinality of that set. $\mathcal{L}_{ij}$ is then the share of words in society i that were borrowed from society j.

At the societal level, we have:

$$\mathcal{L}_i = \sum_{j \in \mathcal{J}} \mathcal{L}_{ij} \tag{2}$$

We define the more general $\mathcal{L}_i$ to simply be the sum of loanwords from each of the various neighbours j. Note that colonial language loanwords would not be included in $\mathcal{L}_i$ since j is limited to the set of geographic neighbours of i.[36] We also drop neighbouring enclaves that speak colonial languages (e.g. we do not have Dutch and indigenous Indonesian languages as neighbours, or English and Xhosa in South Africa).

[36] We do examine colonial languages, but we do so separately in Appendix F.1.

5.B. Constructing Potential Gains from Local Agricultural Trade

To estimate potential gains from local agricultural trade, we use a straightforward Ricardian model with a single factor, in our case agricultural land, and many goods, in our case different crops.[37] We model crop production similarly to Costinot and Donaldson (2012), where production depends on land quantity and the productivity of the land for producing various crops.

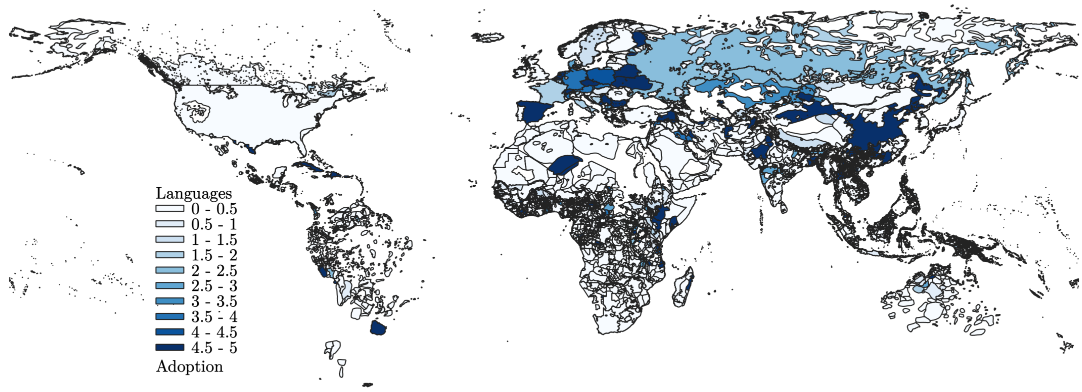
On the demand side, we would ideally have continued to follow Costinot and Donaldson (2012) by assuming price-taking and directly using price data. The differences in the context make this strategy more difficult for us, however. First, Costinot and Donaldson (2012) show that Ricardian trade does a good job of explaining crop production heterogeneity, while we need some notion of utility to measure well-being generated by each neighbour. Second, we narrowly model local trading relationships because it helps our empirical identification to exploit only the plausibly exogenous complementarity of land endowments of a society and its neighbours. This helps to side-step issues like colonialism that likely impacted both trade and language, but not necessarily in a causal manner. We deal with colonialism separately in section 7.[38] One trade-off with this focus is that in very localized trading networks, the price-taking assumption seems less reasonable and we do not observe local prices of all agricultural goods around the world.

For these reasons, we model demand to recover relative prices. We build upon Galor and Özak (2015) by treating societies as having incentive to increase the population they can support, where each adult requires a subsistence bundle of calories and essential nutrients. Galor and Özak (2015) show that caloric potential dominates agricultural suitability. We go one step further by considering the full range of nutritional requirements. This is necessary to be able to consider nutritional complementarity in primary agricultural products, which is known to be an important driver of trade (Gray and Birmingham, 1970).

We compute gains from trade based on a Cobb-Douglas utility function over calories as well as all essential nutrients (see Appendix C for details). We compute the case when a group can trade with all immediate neighbours, and the set of counterfactual cases where we exclude one potential partner at a time for all partners in the group (see again, Figure 2). We explain our methodology for

[37]See Feenstra (2004) for a discussion of this classic model, first introduced in Ricardo (1817).

[38]It helps greatly to avoid considering endogenous heterogeneity in other types of trade. Similarly, we abstract away from trade in goods other than primary agricultural products or globalized trade, to focus on sources of gains from trade where we do not have to worry about the potential endogenous heterogeneity (at the society and good level) in access to long-distance trade routes or productivity in different goods.



(a) Linguistic Adoption

(b) Gains From Trade

Figure 4: Mapping trade incentives and language exchange for Africa

Note: These maps illustrate the main evidence that trade can induce cultural adoption. In panel (a) we map linguistic adoption. Darker shades represent more adoption in those regions. in panel (b) we show gains from trade. Darker shades represent more gains from trade.

estimating trade utility under these counterfactuals in Supplementary Materials section SM-D.4, and report descriptive statistics for our computed gains from trade, and other language group characteristics, in table 1.

i) Validity of the gains from trade measure The basic strategy of using observed relative land productivity to model agricultural trade is valid at the country level (Costinot and Donaldson, 2012). However, since we both model subnational trade, and impose more structure on the demand side, we would ideally like to validate our measure of gains from trade against actual local trade flows in primary agricultural products among subnational language groups. This data, however, does not exist to our knowledge for a large set of groups. We take a number of alternative approaches to validating our measure.

Our main approach is to test whether the model predicts actual production of regionally traded crops, controlling for the FAO productivity of all crops. The

Table 3: Validating the trade measure against actual crop production

	Dependent variable: Actual Land Allocation								
	Sweet potato (1)	Carrot (2)	Sunflower (3)	Sorghum (4)	Coconut (5)	Cassava (6)	Oats (7)	Potato (8)	Global (9)
Model allocation	0.0839***	0.0927***	0.0801***	0.0853**	0.165***	0.225***	0.521***	0.507***	-0.0128
	(0.0230)	(0.0293)	(0.0287)	(0.0391)	(0.0320)	(0.0542)	(0.157)	(0.121)	(0.0959)
Crop suitability	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	2,606	2,606	2,606	2,606	2,606	2,606	2,606	2,606	2,606
R^2	0.247	0.203	0.365	0.380	0.377	0.347	0.375	0.359	0.190
Dep. Var. Mean	0.002	0.0001	0.001	0.007	0.006	0.006	.0002	0.001	0.054

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

crop production data come from Monfreda et al. (2008), who report the share of land allocated to each crop within a 5 arc-minute cell for the whole world.[39]

There are some crops that are typically globally traded, like tobacco, wheat and maize, where contemporary production is clearly not related to the local trade dynamics captured by our model. Accordingly, we define any crop as being predominantly determined by global trade dynamics (and therefore not relevant for our regional trade model) if it had more than five billion USD in global trade in 2008 according to the FAO.[40] We focus more on the crops we expect to be relevant, reporting crop-by-crop estimates for each. We report an aggregate for the crops that are predominantly traded globally, which serves as a placebo estimate. For each, we regress the actual production on the model estimated production, and the vector of suitabilities of all FAO crops. We run the regression at the society-level (denoted i), resulting in the following regression equation:

$$(3)\quad ActualProduction_i = \beta_0 + \beta_1 PredictedProduction_i + \Gamma FAOSuitability_i + \epsilon_i$$

We are interested in β_1 and report estimates of this parameter in Table 3. The estimates for crops that we believe to be relevant for our trade model are in columns 1-8. Each one has (precisely estimated) predicted production that is positively associated with actual production. For the globally traded crops, our model is not relevant, as anticipated (column 9).

Our second approach is to look at historical market prices for crops across a

[39] We focus on the society mean of that variable for each crop with more than 0.0001% mean land share for both actual and estimated mean land allocation, that exist in both the FAO and the Monfreda et al. (2008) dataset.

[40] This delineation was based on a a natural gap in export dollars that exists in the data (see figure E1), and results in 14 crops in the global trade group and 8 crops in the group we expect to be relevant for our model.

number of cities, sourced from Jacks (2004, 2005). We matched these cities to our language groups to test whether our model-predicted gains from trade between a pair of neighbouring groups is correlated with market price integration.[41] This gives us only sixty pairs where we can match to our data, but on this small sample we show in table SM-C1 that model-predicted gains from trade are associated with greater price integration, suggesting greater trade volume among these pairs.[42]

ii) From the trade model to hypothesis testing: variable definition All of this comes together into a few variables as follows. At the society-pair level, where U_i^{FT} is society i's utility under full trade, and U_i^{FT-j} is society i's utility under full trade without society j, as in Figure 2, we define:

$$c_{ij} = \frac{U_i^{FT} - U_i^{FT-j}}{U_i^{FT-j}} \tag{4}$$

Which specifies the contribution of j to the trade utility of i. Note that $c_{ij} < 0$ if j is a competitor to i and $c_{ij} > 0$ if i and j make natural trade partners from the perspective of i. We plot a histogram of c_{ij} in figure E2(a), and find that it is centred approximately around zero.

Note that in the figure, we show only the range {-10, 10}, to avoid severe x-axis distortion.[43] The maximum value of the variable is over 3,000, and there are actually more than a handful of society-pairs with values over 100.[44] To avoid estimates that are driven by societies in the tail of the distribution, and that may have undue influence on a regression we show results with the percent change variable, as in equation 4,[45] and a percentile rank version of the same variable.[46]

At the society-level - guided by our hypothesis in section 3 - we focus on the neighbour with the biggest gains from trade:

$$c_i = \max_j \{\frac{U_i^{FT} - U_i^{FT-j}}{U_i^{FT-j}}\} \tag{5}$$

[41] See Supplementary Materials section SM-C.1 for details on this data and the matching procedure.

[42] In Table SM-D1 we conduct a third validation exercise based on population that further reinforces the validity of our trade model.

[43] We do this only for the purpose of the visualization, not the broader analysis. This eliminates about 2% of the sample, typically ones with very large values rather than very small. Even a value of 10 is very large, it suggests a single neighbour improves the welfare of a society by 10-fold.

[44] These are typically societies that would not be able to survive without a particular neighbour, and thus have a near-zero denominator.

[45] Here we deal with the outliers by winsorizing at 5%.

[46] The percentile rank is obviously uniformly distributed, and is between 0 and 1.

Note that $c_i = 0$ if i is, at best, indifferent towards trade, and in practice most societies have at least one partner that they are at least indifferent towards. $c_i > 0$ is therefore typical, which can be seen in figure E2(b), where we plot the c_i histogram.[47] As with the pairwise measure, we show both percent change (as in equation 5) and a percentile rank version of the same variable.

In addition to exploring how much a society's gains from trade is related to language adoption, our hypothesis predicts that incentives to interact also drive lending of loanwords to neighbours. We therefore construct a measure of the influence of society i on the agricultural trade of neighbour j. This is constructed analogously to the adoption measures, and we denote the resulting influence variables ι_{ij} and ι_i at the relationship and society levels, respectively.

6. Results: trade-incentives and linguistic convergence

Our main focus is to empirically investigate the extent and causes of linguistic convergence. We begin by noting a negative correlation between language adoption and influence (figure 5). Notably, if loanwords were purely determined by contact the correlation would be positive. Can loanwords be thought of as the result of decisions by individuals in a society to invest in trading relationships?

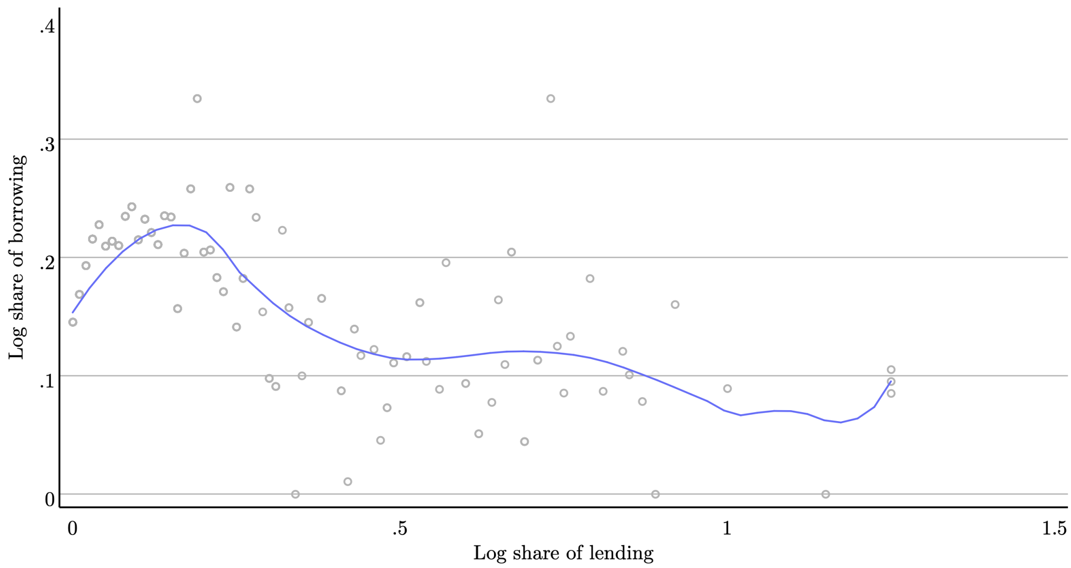


Figure 5: Correlation between adoption and lending

Note: The figure shows the correlation between language adoption and language lending. The figure plots $log(1 + \mathcal{L}_i)$ (where $\mathcal{L}_i$ is as defined in equation 2) as well as the analogous measure for lending. Both of these measures are then winsorized at the 0.1% level to regulate the axis-scale. The scatterplot groups observations into 0.01 lending bins. The fit line is based on a biweight kernal of degree 1, with a bandwidth of 0.025.

[47]We also show, in Figure 4(b) how this measure is distributed spatially.

6.A. Prediction 1: Local agricultural trade incentives and language exchange

Our first prediction is that we should expect a positive correlation between the gains from trade with a society's best trading partner and the total loanwords in the society's language. Recall that only the gains from trade with a society's best neighbour *unambiguously* increases linguistic borrowing, since gains from trade with other societies can, in theory, change the distribution of bilingual speakers in a way that reduces diffusion.[48]

In figure 6 we see the positive correlation between gains from agricultural trade and linguistic exchange that we expect. What is more surprising though, is that language exchange in *both* directions is strongly associated with a society's gains from agricultural trade, as well as their influence on local agricultural trade. We would have expected a correspondence only between linguistic borrowing and trade gains, and linguistic lending and trade influence. In other words, in figure 6, we expected a positive correlation in subfigures (a) and (d), but not (b) and (c). So, the figure at first glance appears inconsistent with our predictions.[49]

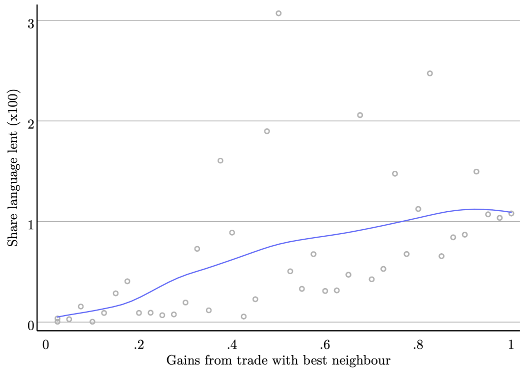
This suggests a puzzle: figure 6 represents a reversal from figure 5, which as noted above *is* consistent with our model. This apparent contradiction might be explained by the extremely strong correlation between gains from trade and trade influence (figure E5), which indicates that some of these raw correlations may be spurious. Accordingly, we re-investigate using the following horse-race style regression between gains from trade and trade influence:

$$\mathcal{L}_i = \alpha_{colonizer} + \alpha_{continent} + \beta_1 c_i + \beta_2 \iota_i + X_i'\Gamma + \epsilon_i \tag{6}$$

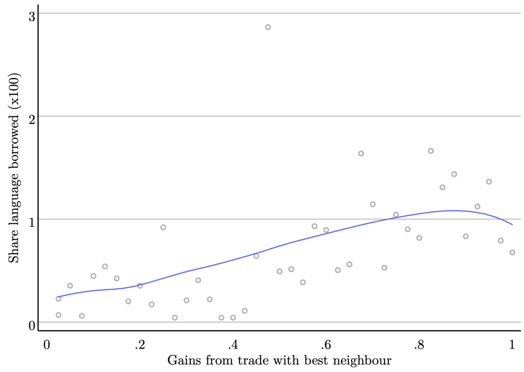
The outcome variable $\mathcal{L}_i$ represents the total share of words in a society (denoted i) borrowed from neighbours. This is defined formally in equation 2. We include fixed effects for colonizer and continent, denoted $\alpha_{colonizer}$ and $\alpha_{continent}$ respectively. X' is a vector of controls, including the following: agricultural wealth (structurally estimated); the amount of arable land; and a quintic polynomial in average distance to neighbours. The contact hypothesis predicts that language convergence is a function of the population of the society, its neighbours, and their ratio, so we also include each of those in X'. We also include two measures of land

[48] Beyond the conceptual motivation for focussing on the best neighbour, we should note that the best neighbour is a good proxy for neighbourhood quality (figure SM-F1). That is, the regions with the best best neighbours also have the best worst neighbours.

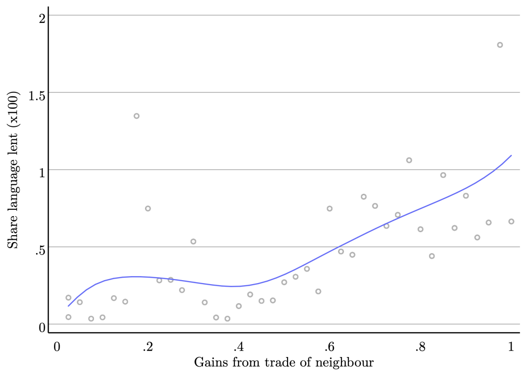
[49] In fact it appears more consistent with a simple model of passive, contact-based language exchange (the contact hypothesis). This is the model that is typically discussed in the socio-linguistics literature.



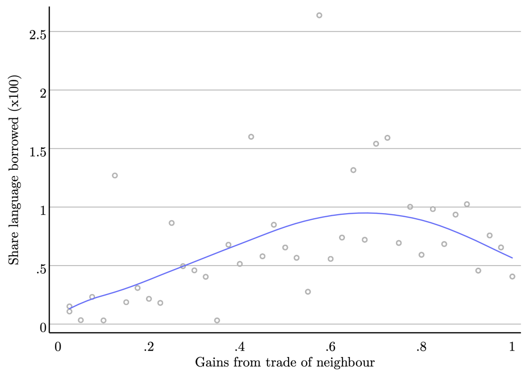
(a) Lang. adoption and gain from trade



(b) Lang. influence and gain from trade



(c) Lang. adoption and trade influence



(d) Lang. influence and trade influence

Figure 6: Gains from trade and language exchange

Note: The figure shows the correlation between language exchange and trade gains/influence. The figure plots $\mathcal{L}_i$ (where $\mathcal{L}_i$ is as defined in equation 2) as well as the analogous measure for lending. Both of these measures are then winsorized at the 1% level to regulate the axis-scale (this is not done in the analogous tables). The scatterplot groups observations into 0.01 lending bins. The fit line in each graph is based on a biweight kernal of degree 1, with a bandwidth of 0.035.

diversity to account for the ethnolinguistic convergence mechanism proposed by Michalopoulos (2012).[50] The two variables we are interested in are c_i and ι_i. c_i is the gains from trade variable, defined in equations 5, while ι_i is the analogous trade influence measure.

In this framework, we expect $\beta_1 > 0$; $\beta_2 = 0$. We also examine linguistic lending as an outcome. In that case the only difference is that we expect $\beta_1 = 0$; $\beta_2 > 0$. The results are in table 4. As in figure 6, gains from trade are associated with linguistic borrowing, and estimates using the percentage of welfare gain (columns 1-2) and the percentile rank of gains from trade (columns 3-4) are consistent with each other. Similarly, trade influence is associated with linguistic lending using either the percentage welfare measure (columns 5-6) or the percentile

[50] The first is the mean variance in suitability of each crop, and the second takes the sum of the absolute difference in crop suitability for all crops.

rank (columns 7-8). So, the table highlights the robustness of the correlations in figure 6 (a) and (d) to the various controls and fixed effects in the regression. However, perhaps more importantly, gains from trade remains strongly correlated with borrowing controlling for influence (columns 2 and 4), and vice versa (columns 6 and 8).[51]

The percentage change measure may be the more helpful to interpret the magnitudes (0.085 in column 2 and 0.29 in column 6). On average, a society's best neighbour improves their welfare by about 96% (table 1), which corresponds to a roughly 0.09 percentage point increase in language adoption. This means that gains from trade with a typical society's best trade partner contributes to about 10% of the regional loanword adoption of a typical society. If we look at the percentile rank measure, going from the bottom to the top of the distribution represents about 1 percentage point change in borrowing (columns 3 & 4), which is about the mean of loanword adoption. Keeping in mind that the estimates capture only local trade in agricultural goods, these estimates seem plausible and appropriate.

Also important is that once we control for gains from trade, trade influence no longer positively covaries with language exchange (columns 2 and 4). In column 2 the estimate is actually negative, while in column 4 it is nearly an order of magnitude smaller than the gains from trade estimate. This null result is a first step to distinguish between our model and the contact hypothesis (where we might expect *both* gains and influence to covary with loanwords). Similarly in columns 6 and 8, we find that the effect of gains from trade is much less important than trade influence for language lending.

Our conceptual framework guided our decision to focus on the gains from trade with the best neighbour and total language borrowing. However, we do have clear predictions both about the implications for borrowing from only the best neighbour, as well as if we try the - perhaps *a priori* more obvious - specification regressing average language adoption on average gains from trade. For the first, we show, as anticipated, a nearly identical set of results when we look at the gains from trade with the best neighbour and adoption from the best neighbour (table E2, columns 1-2). We also report, in columns 3 and 4 the estimates based on gains from trade with an average neighbour and average borrowing. In this case, as expected, we see much weaker results - over six times smaller - albeit in the same direction. The weaker results are expected because now new adoption can increase the number of bilingual speakers, but also crowd out bilingualism of other

[51]Robustness to different rules for determining what a loanword is appears in table E1.

Table 4: Gains/influence from agricultural trade and language borrowing/lending

Dependent Variable:	Language Borrowed				Language Loaned			
Utility measure	percent change		percentile rank		percent change		percentile rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gains from trade with neighbours	0.0776**	0.0854**	0.901***	0.851***		0.0521		0.234
	(0.0337)	(0.0353)	(0.198)	(0.298)		(0.0753)		(0.322)
Influence on trade with neighbours		-0.0328		0.102	0.309***	0.294**	1.875***	1.743***
		(0.0368)		(0.322)	(0.115)	(0.129)	(0.464)	(0.554)
Trade wealth (structurally estimated)	✓	✓	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓	✓	✓
Land Share	✓	✓	✓	✓	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓	✓	✓	✓
N	2,602	2,602	2,602	2,602	2,602	2,602	2,602	2,602
R^2	0.064	0.064	0.067	0.067	0.120	0.121	0.121	0.121
Dependent Variable Mean	0.938	0.938	0.938	0.938	0.938	0.938	0.938	0.938

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the c_i measure defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, in order to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours.

languages. In columns 5 and 6 we show negative estimates on gains from trade with the best neighbour and language exchange with the worst, which highlights the crowding-out that was driving the smaller estimates in columns 3 and 4, and which was hypothesized in section 3.

One consideration that could be of interest is how recent is the cultural convergence. A society that is concerned with cultural convergence may be far less concerned if it takes thousands of years to materialize. To determine whether the data is capturing dynamics over the long run, short run, or both, we estimate our gains from trade model using only crops that would have been available to a society prior to the Colombian Exchange, and separately, crops that would have been available only after.[52] The Colombian Exchange refers to the widespread exchange of agricultural products between Old and New World around the 15th and 16th centuries. So, if trade of Old World crops in the New World predicts loanword borrowing then this dates linguistic borrowing to sometime after the 15th century. These estimates are available in table E3. We estimate effects for the post-Colombian exchange crops (column 1) that are about equal to that of the pre-Colombian exchange crops (columns 2), so our main estimates may to be capturing a combination of persistent historical changes as well as more recent ones.[53] Even in horse-race specifications, both are positively associated with

[52] We distinguish between New and Old world as in Nunn and Qian (2011).

[53] Note that the pattern we find is consistent with all adoption being recent, but we cannot

loanwords and remain similar in magnitude (columns 3-4).

A final consideration is about identification. How do we know that economic trade is the reason why trade incentives are associated with language exchange? There could, after all, be some unobserved variable that is correlated with both gains from trade and language exchange. One way to investigate this possibility is to examine unviable trading relationships. For these society-pairs we can still observe a continuous gains from trade variable, however we do not expect that this variable will impact actual trade for these societies. Once a relationship is not viable there will not be trade regardless of how unviable it is. We show this falsification test in table E4. Indeed, we find no evidence that gains from trade influences language exchange for unviable trading relationships, reinforcing our belief that our main results are in fact driven by local agricultural trade.

We conclude that gains from trade can be associated with language exchange in the expected direction, so prediction 1 has strong support in the data.

6.B. Prediction 2: Adoption is inverse-U shaped in the number of partners

As outlined in Section 3, trading with a new partner can attract bilingual speakers who might otherwise have been unilingual, or crowd-out those that may have otherwise learned another language. These two competing effects - the increase in bilingualism and the reduced intensity of bilingualism in any particular language - are expected to generate an inverse-U shape in borrowing.[54,55] On the other hand, when a society influences many neighbours, this does not necessarily influence the intensity of bilingualism in any language.[56] In that case we expect only the positive of the two effects, so we hypothesize a weakly increasing relationship for lending.

Since we do not observe the crowd-out rates[57] we proxy for diversity with the number of viable trading neighbours a society has. Figure 7 presents the raw data. It shows nearly exactly the theoretical prediction of an inverse-U in borrowing and weakly increasing relationship in lending. We can examine this prediction in a regression framework as well. We compare linear and quadratic

empirically rule-out a combination of recent and historical adoption.

[54]See section 3 for a more precise description

[55]Recall that additional lending does come with some substitution. This was previously highlighted when we described the results presented in table E2.

[56]We focus on viable neighbours, where at least one of the two parties must find trade profitable for the partnership to be viable. We use the same definition of viability here as we did for the results presented in table 6.

[57]σ_{jV} from the theory

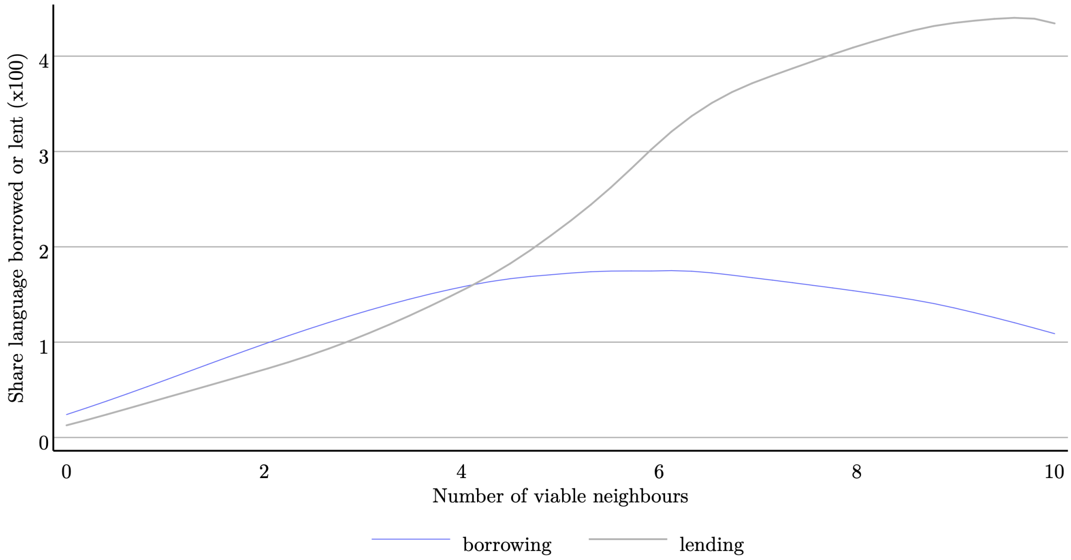


Figure 7: Viable trade partners and language borrowing

Note: The figure shows the relationships between language exchange and the number of viable trading relationships a society has. Our hypothesis is that adoption is inverse-U shaped in number of neighbours while lending is increasing. A viable trading relationship is defined as any society-pair where at least one of the two societies may gain from trade. The fit lines are based on bi-weight kernels of degree 1 with a bandwidth of 4. Both adoption and lending are winsorized at 1%.

empirical specifications in the number of viable neighbours. The controls are all the same as in our main society-level regression (equation 6).

We focus on viable economic relationships in the regressions. We do this because the mechanism specified by our conceptual framework is that individuals choose to invest in languages to facilitate trade, so we do not expect reductions in losses from trade to be meaningful. The relationships that did not generate gains would not come to fruition in the first place. Just as in the theory, we define a viable relationship as one where there exist positive gains from trade. We also control for the total number of neighbours in our main specification to ensure that we are capturing the effect of viable neighbours. The results are robust to not including this control however, and this can be seen in table E5.[58]

The estimates for both borrowing and lending are in table 5. In column 1 we test the linear specification and in column 2 we examine the quadratic, both for linguistic borrowing. We find that the quadratic model fits the data better; in column 2 both the linear and quadratic terms are significant, and suggest the inverse-U pattern that we hypothesized. If we take the parameter estimates seriously, they suggest that linguistic borrowing peaks between 4 and 5 neighbours -

[58]We also look at robustness to an alternate construction of 'viability' of neighbours in table SM-F1. again, the results are unaffected.

about 16% of societies have 4 or more viable neighbours.[59]

Table 5: Total adoption and the supply of viable trade partners

Dependent Variable:	Language Borrowed		Language Loaned	
	(1)	(2)	(3)	(4)
Number of viable trading neighbours	-0.00392	0.192**	0.173	-0.0278
	(0.106)	(0.0948)	(0.231)	(0.309)
Number of viable trading neighbours squared		-0.0196***		0.0201
		(0.00559)		(0.0463)
Total Neighbours	0.102	0.117	0.314**	0.298**
	(0.0952)	(0.0969)	(0.126)	(0.134)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
N	2,601	2,601	2,601	2,601
R^2	0.065	0.074	0.125	0.131
Dependent Variable Mean	0.936	0.936	0.932	0.932

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Viable trading relationships are any relationships where at least one of the two parties can gain from trade. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours.

For lending we do not find the same. In contrast with linguistic borrowing, the quadratic term in the lending regression in column 4 is not statistically significant. The point estimates do not even go in the same direction in the quadratic model, so we cannot compute the same peak.

6.C. Prediction 3: Language exchange and asymmetric gains from trade

Prediction 3 was that the society within a relationship that gains the most from trade will be the one to bear the fixed cost of adoption (see section 3). Investigating this prediction empirically requires us to move from a society-level analysis to a relationship-level analysis. We can now include relationship fixed-effects to look within a relationship and assess whether the society that benefits more, borrows more (or equivalently the one who benefits less, lends more).

[59] In column 2 the linear term is about 0.19, the quadratic is about -0.02. This suggests that the number of partners at the peak of the inverse-U (p^*) is $0.19 - 2 \cdot 0.02p^* = 0$; $p^* = 4.75$

The introduction of the relationship fixed-effects additionally allows us to control for much more than we previously could, and serves as a robustness check to the analysis previously discussed. Adding these fixed effects precludes us from being able to independently examine gains from trade and trade influence, but we can still check this at the relationship-level in a specification that includes fixed-effects for each group separately. Accordingly, at the relationship level we test specifications both with relationship-level fixed effects, as well as ones with fixed effects for each society. These are:

$$\mathcal{L}_{ij} = \alpha_{ij} + \beta_1 c_{ij} + X'_{ij}\Gamma + \epsilon_{ij} \tag{7}$$

$$\mathcal{L}_{ij} = \alpha_i + \alpha_j + \beta_1 c_{ij} + \beta_2 c_{ji} + X'_{ij}\Gamma + \epsilon_{ij} \tag{8}$$

Everything is as previously described, with the exception of the subscript ij which denotes either the loanwords adopted by society i from society j, or in the case of trade, ij denotes the gains of i by trading with j. We also omit some of the controls that we had in the society-level analysis since they are made redundant by the inclusion of the relationship fixed-effects (α_{ij}) or the society fixed-effects (α_i and α_j). As before we expect $\beta_1 > 0$ and $\beta_2 = 0$.

The estimates generated by these regressions can be seen in table 6.[60] The two models generate results that are both consistent with each other, and with the results presented in table 4. The estimate from the relationship fixed-effects model (column 1) suggests that if a society gains 10% more from trade than their partner, they borrow about double the typical loanwords of a viable relationship. So this implies that this 10% gains from trade is enough to have one party take on all of the language adoption in a typical viable relationship. This suggests that societies do borrow more when relationships are more profitable, for the same level of contact (whenever i interacts with j, j necessarily interacts with i). This finding highlights in the clearest way so far, that cultural exchange is based on an optimization decision rather than an unintentional byproduct of contact. Investigating the relationship using the percentile rank - which is less prone to influence from outliers generates no change in interpretation (column 2), and the specifications with the individual group fixed-effects, rather than the relationship fixed effects are also consistent with all previous findings (columns 3 & 4).

We can also investigate - as we did with the societal-level analysis - whether

[60] Robustness to various loanwords thresholds is in table E6.

Table 6: Loanwords and trade incentives at the relationship level

Dependent Variable:	Language Borrowed			
Utility measure:	percent change	percentile rank		
	(1)	(2)	(3)	(4)
Gains from trade with neighbours	0.0646**	0.649**	0.403**	0.416**
	(0.0281)	(0.262)	(0.198)	(0.199)
Influence on trade with neighbours				0.462
				(0.337)
Relationship Fixed Effects	✓	✓		
Society Fixed Effects (both)			✓	✓
Baseline controls	✓	✓	✓	✓
N	5,561	5,561	5,561	5,561
R^2	0.521	0.522	0.661	0.661
Dependent Variable Mean	0.277	0.277	0.277	0.277

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the c_i measure defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case we aggregate to the society level by taking the maximum value from the society's neighbours. Viable trading relationships are any relationships where at least one of the two parties can gain from trade. Language Borrowed (range [0,100]) is defined in equation 1. 'Society Fixed Effects (both)' means we separately include a society-fixed effect for each society in the relationship; 'Relationship Fixed Effects' means we include a fixed effect for a specific pair. Controls are as follows: trade wealth (estimated); population; land share; land diversity.

this is based on long or short run dynamics by examining the Colombian Exchange. Table E7 shows effects for recent and pre-Colombian exchange crops that are about equal, as they were at the societal-level. We can also investigate the unviable relationships (table E8). Columns 1 and 2 both highlight a null-effect, which improves our confidence in the empirical strategy that we have adopted.

6.D. Is language adoption cultural?

One remaining issue regarding linguistic borrowing is the nature of the exchange. The main question posed in the paper relates to cultural convergence, and we have so far assumed that loanwords reflect a reduction in cultural distance. However, another reasonable interpretation could be that language exchange is purely functional.[61] For instance, a society benefiting greatly from being able to trade for broccoli may benefit precisely because they do not have any other source for broccoli, through either production or trade. As a result, their only exposure to any word for broccoli would come through the trading partner. In this case borrowing

[61] For ease of discussion we will distinguish between these mechanisms using the labels 'cultural borrowing' and 'functional borrowing.'

the word would not reflect any investment in bilingualism, it would simply reflect the functional adoption of words for lack of any alternative.

One nice feature of our data is that it exists at the word-level, which allows us to get very precise about the types of words that get borrowed. We can therefore test directly whether the linguistic borrowing that is driving our results is functional or cultural. The most direct test of functional borrowing would be to see if the words driving the borrowing are the words for the specific crops that are being exchanged, or other more general words. To capture this we took the list of names of our crops in the trade data, and implemented a semantic-analysis routine to get the various representations of these crop names in each language (see appendix D for details). We then re-run the borrowing analysis for only these crop names, as well as for all words *except* the crop names. We run a similar procedure for words relating to economic transactions.[62]

Table 7: Loanwords by word-type and trade incentives

	Crop names	Non-crop words	Economic transaction words	All but crop/transaction
	(1)	(2)	(3)	(4)
Gains from trade with neighbours	0.163*	0.461***	0.0857**	0.441***
	(0.0874)	(0.147)	(0.0349)	(0.149)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
N	2,499	2,499	2,499	2,499
R^2	0.027	0.096	0.023	0.097
Dependent Variable Mean	0.257	0.555	0.107	0.523

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We lose 90 observations relative to the sample in table 4 because there are some languages where we find no english equivalents for one category of word-type, and these observations are dropped across all specifications to facilitate comparisons within the table. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, which is aggregated to the society level by taking the sum of the society's neighbours. All word-type adoption outcomes are winsorized at the 0.1% level to deal with outliers. See table D1 for the wordlists used to generate these classifications.

The heterogeneity by word-type is presented in table 7. Columns 1 and 2 show the analysis for crop-names and non-crop words respectively. We find evidence of both cultural and functional borrowing. Column 1 shows borrowing for crop names, while column 2 shows non-crop words. Both estimates are positive, but

[62] Again, appendix D includes details of how this was constructed and table D1 for the origin wordlists used.

it is difficult to compare magnitudes since the set of non-crop words is much larger (figure SM-F2). Column 3 examines words for economic transactions since functional borrowing could include concepts like 'money,' or 'contract.' Column 4 shows all other words - those we believe unrelated directly to the economic exchange captured by our trade incentives variable, and therefore more likely to reflect cultural adoption. Again, we see precise, positive estimates for both functional and non-functional word-types.

However, for some, the table may not directly address *cultural* adoption since not all non-functional words are necessarily cultural. Our main interest here is to show that our results are not exclusively driven by functional words, so columns 2 and 4 of table 7 simply remove these words to demonstrate similar estimates. But the word-types that are driving estimates may be of interest as well. Without taking a stance on where to draw the line between functional and cultural words (or whether such a line is even conceptually appropriate), in table E9 we investigate a number of word categories: politics, religion, human rights, social organization, and technology.[63] In each case we see more borrowing. Interestingly increased adoption of religious words is the most precise category of any word type, cultural or functional. Given that each of political structure, religion, and human rights produce positive, precise estimates, we are confident that the borrowing we capture is at least partly, even if not fully cultural.

We discuss another approach to examine cultural borrowing in Supplementary Materials section SM-E.1, and based on that empirical exercise we come to the same conclusion.

7. Additional Results

7.A. Colonialism

We investigate colonialism, which had a very clear role in shaping language across the globe. In appendix F.1 we highlight two things. As expected, colonial presence is associated with more colonial language adoption. This should be expected on the basis of the huge power asymmetry and associated coercion. However we also find that colonial intensity appears unrelated to local gains from trade and trade influence. Further, those that borrowed more of the colonial language, were not more or less likely to borrow or lend to their local neighbours. Accordingly it appears unlikely that colonialism is the mechanism behind our core results.

[63] Details on how these measures were constructed are in appendix D.

7.B. *Diversity*

Second, we consider diversity more broadly. We are interested in cultural convergence, but we might think of fractionalization as cultural convergence (or a lack thereof) at its limit. This then begs the question: do the same trade incentives generate ethnolinguistic homogenization? We show in appendix F.2 that indeed the same trade incentives that generate loanword exchange in our data also generate reductions in country-level diversity, as measured by each of the main diversity measures in the literature.

8. Conclusion

In this paper we construct a large dataset that allows us to directly observe cultural adoption globally. This dataset features not only information on adoption, but it also precisely measures the source and direction of exchange between society-pairs. We also develop a methodological approach to investigate within-country gains from economic trade.

We find that while trade incentives do cause cultural convergence, the context matters a lot. We generated three theoretical predictions that we test empirically. The first is that gains from trade, especially with a society's best trading partner generates cultural convergence. This prediction was strongly supported by the data, as gains from trade with a society's best neighbour increases total loanword adoption; gains from trade with an average neighbour increases adoption by much less; while increased gains from trade reduced total loanword adoption from poor trading partners. The last result is likely due to the fact that the increased adoption of on language can crowd-out bilingualism in other languages.

This is the basis of our second theoretical prediction, which is that there is an inverse-U shape in loanwords adoption for the number of viable trade partners a society has. In the data we can confirm this as well, there is a strong quadratic relationship that peaks between 4 and 5 viable neighbours - this quadratic relationship appears in the data for borrowing but not lending. Finally, theoretically we expect strong asymmetries in borrowing within relationships. This final prediction distinguishes models of cultural convergence based on optimization and purposeful decisions from models in the socio-linguistics literature that view cultural convergence as an unintended byproduct of contact. We find that borrowing is much heavier for societies that benefit more from trade. In fact, just a 10 percentage point gap in gains from trade is enough to tilt all cultural convergence within a typical viable trading relationship to one of the two parties.

We conclude that there is a role for cultural protection policies in societies with low trade leverage that are concerned with cultural erosion, albeit quite a limited one. For instance, our results suggest, perhaps surprisingly, that cultural protectionism can *strengthen* the homogenizing effect of trade on culture if it leads to fewer trade partners, with higher intensity. Furthermore, protectionist policies can be unnecessary sticking points in trade negotiations, if they are intended to protect the culture of a society that already has trade leverage. All of this suggests that some of the (significant) concerns regarding the effects of free trade on culture may not be completely unfounded, but may in many cases be somewhat overstated.

References

Pelle Ahlerup and Ola Olsson. The roots of ethnic diversity. *Journal of Economic Growth*, 17(2):71–102, 2012.

Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic growth*, 8(2):155–194, 2003a.

Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic Growth*, 8:155–94, 2003b.

Alberto Alesina, Paola Giuliano, and Nathan Nunn. On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2): 469–530, 2013.

Alberto Alesina, Stelios Michalopoulos, and Elias Papaioannou. Ethnic inequality. *Journal of Political Economy*, 124(2):428–488, 2016.

Yann Algan and Pierre Cahuc. Inherited trust and growth. *American Economic Review*, 100(5):2060–92, 2010.

Yann Algan, Thierry Mayer, Mathias Thoenig, et al. The economic incentives of cultural transmission: Spatial: Spatial evidence from naming patterns across france. Technical report, 2013.

Quamrul Ashraf and Oded Galor. Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, 103(3):528–33, 2013a.

Quamrul Ashraf and Oded Galor. The ‘out of africa’ hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46, 2013b.

David Atkin. The caloric costs of culture: Evidence from indian migrants. *American Economic Review*, 106(4):1144–81, April 2016.

David Atkin, Eve Colson-Sihra, and Moses Shayo. How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy*, forthcoming.

Michele Belluzzi. Cultural protection as a rationale for legislation: the french language law of 1994 and the european trend toward integration in the face of increasing us influence. *Dick. J. Int’l L.*, 14:127, 1995.

Alberto Bisin and Thierry Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory*, 97(2):298–319, 2001.

Alberto Bisin and Thierry Verdier. Trade and cultural diversity. In *Handbook of the Economics of Art and Culture*, volume 2, pages 439–484. Elsevier, 2014.

Leonard Bloomfield. Language, 1933.

Arthur Blouin. Culture and contracts: The historical legacy of forced labour. *University of Toronto Working Paper*, 2020.

Arthur Blouin and Sharun W Mukand. Erasing ethnicity? propaganda, nation building, and identity in rwanda. *Journal of Political Economy*, 127(3):1008–1062, 2019.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

Jeremy Bulow and Paul Klemperer. The Generalized War of Attrition. *The American Economic Review*, 89(1):175–189, 1999. Publisher: American Economic Association.

Deborah Cameron and Don Kulick. *Language and sexuality.* Cambridge University Press, 2003.

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

J X Chipponi, J C Bleier, M T Santi, and D Rudman. Deficiencies of essential and conditionally essential nutrients. *The American Journal of Clinical Nutrition*, 35(5):1112–1116, 1982.

Arnaud Costinot and Dave Donaldson. Ricardo's theory of comparative advantage: old idea, new evidence. *American Economic Review*, 102(3):453–58, 2012.

Patricia Cukor-Avila and Guy Bailey. The effects of the race of the interviewer on sociolinguistic fieldwork. *Journal of Sociolinguistics*, 5(2):252–270, 2001.

Partha Dasgupta and Ismail Serageldin. *Social capital: a multifaceted perspective*. The World Bank, 1999.

Roy Davies. Current Value of Old Money. URL http://projects.exeter.ac.uk/RDavies/arian/current/howmuch.html.

Anthony Depalma. 19 nations see u.s. as threat to cultures. *New York Times*, July 1, 1998.

Klaus Desmet and Romain Wacziarg. The cultural divide. *National Bureau of Economic Research Working Paper Series*, No. 24630, 2018.

Klaus Desmet, Shlomo Weber, and Ignacio Ortuño-Ortín. Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7(6):1291–1318, 2009.

Andrew Dickens. The Historical Roots of Ethnic Differences: The Role of Geography and Trade. Working Paper 1901, Brock University, Department of Economics, June 2019.

Steven N. Durlauf and Marcel Fafchamps. *Social Capital*. Amsterdam: North Holland, 2005.

Joan Esteban, Laura Mayoral, and Debraj Ray. Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–42, 2012.

Pablo D Fajgelbaum, Pinelopi K Goldberg, Patrick J Kennedy, and Amit K Khandelwal. The return to protectionism. *The Quarterly Journal of Economics*, 135 (1):1–55, 2020.

FAO. *FAO/INFOODS Food Composition Database for Biodiversity Version 4.0, BioFoodComp4.0.* FAO. Rome, Italy, 2017a.

FAO. *FAO/INFOODS Analytical food composition database version 2.0, AnFooD2.0.* FAO. Rome, Italy, 2017b.

Robert C. Feenstra. *Advanced international trade: theory and evidence.* Princeton University Press, Princeton, N.J, 2004.

Gabriel J Felbermayr and Farid Toubal. Cultural proximity and trade. *European Economic Review*, 54(2):279–293, 2010.

P. Frankopan. *The Silk Roads: A New History of the World.* Knopf Doubleday Publishing Group, 2016.

Jean Gabszewicz, Victor Ginsburgh, and Shlomo Weber. Bilingualism and communicative benefits. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, pages 271–286, 2011.

Oded Galor and Ömer Özak. Land productivity and economic development: Caloric suitability vs. agricultural suitability. *Working Papers - Brown University, Department of Economics*, (2015-5), 2015.

Oded Galor and Ömer Özak. The agricultural origins of time preference. *American Economic Review*, 106(10):3064–3103, 2016.

Jeffrey Garten. Capital imperialism is no joke. *Business Week*, 1998.

Paola Giuliano and Nathan Nunn. Ancestral characteristics of modern populations. *Economic History of Developing Regions*, 33(1):1–17, 2018.

Paola Giuliano, Antonio Spilimbergo, and Giovanni Tonon. Genetic distance, transportation costs, and trade. *Journal of Economic Geography*, 14(1):179–198, 2014.

Edward L Glaeser, David Laibson, and Bruce Sacerdote. An economic approach to social capital. *The economic journal*, 112(483):F437–F458, 2002.

Gunes Gokmen. Clash of civilizations and the impact of cultural differences on trade. *Journal of Development Economics*, 127:449–458, 2017.

R. Gray and D. Birmingham. *Pre-Colonial African Trade: essays on trade in Central and Eastern Africa before 1900.* Oxford U.P., 1970.

Gene M Grossman and Elhanan Helpman. Identity politics and trade policy. 2020.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Social capital as good culture. *Journal of the European Economic Association*, 6(2-3):295–320, 2008.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Cultural biases in economic exchange? *The quarterly journal of economics*, 124(3):1095–1131, 2009.

John J Gumperz and Robert Wilson. Convergence and creolization. *Pidginization and creolization of languages*, 151:167, 1971.

Robert E Hall and Charles I Jones. Why do some countries produce so much more output per worker than others? *The quarterly journal of economics*, 114 (1):83–116, 1999.

M. Haspelmath and U. Tadmor. *Loanwords in the World's Languages: A Comparative Handbook*. De Gruyter Mouton, 2009a.

Martin Haspelmath and Uri Tadmor, editors. *WOLD*. Max Planck Institute for Evolutionary Anthropology, 2009b. URL https://wold.clld.org/.

Jeffrey G Heath. Language contact and language change. *Annual review of anthropology*, 13(1):367–384, 1984.

Paul R. Hensel and Sara M. Mitchell. The Issue Correlates of War (ICOW) Project Supplementary Data Set: Colonial History Data Set, 2007. URL https://doi.org/10.7910/DVN/5EMETG. Edition: V2 Section: 2007-11-28 00:00:00.0.

Bates L Hoffer. Language borrowing and language diffusion: An overview. *Intercultural Communication Studies*, 11(4):1–37, 2002.

IIASA/FAO. *Global Agroecological Zones (GAEZ v3.0)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy, 2012.

Institute of Medicine. *Dietary Reference Intakes: The Essential Guide to Nutrient Requirements*. The National Academies Press, 2006.

David S Jacks. Market Integration in the North and Baltic Seas, 1500-1800. *The Journal of European Economic History*, 33(2), 2004.

David S. Jacks. Intra- and international commodity market integration in the Atlantic economy, 1800–1913. *Explorations in Economic History*, 42(3):381–413, 2005. Publisher: Elsevier.

Matthew A. Jaro. Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. *Journal of the American Statistical Association*, 84(406):414–420, 1989.

Saumitra Jha. Trade, institutions, and ethnic tolerance: Evidence from south asia. *American political Science review*, pages 806–832, 2013.

Saumitra Jha and Moses Shayo. Valuing peace: the effects of financial market exposure on votes and political attitudes. *Econometrica*, 87(5):1561–1588, 2019.

Karsie A Kish. Protectionism to promote culture: South korea and japan, a case study. *U. Pa. J. Int'l Econ. L.*, 22:153, 2001.

István Kónya. Modeling cultural barriers in international trade. *Review of International Economics*, 14(3):494–507, 2006.

Paul Krugman. Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499, 1991. Publisher: University of Chicago Press.

William Labov. Phonological correlates of social stratification. *American Anthropologist*, 66(6):164–176, 1964.

William Labov and Wendell A Harris. Addressing social issues through linguistic evidence. *Language and the law*, pages 265–305, 1994.

M. Lalee, J. Nocedal, and T. Plantenga. On the implementation of an algorithm for large-scale equality constrained optimization. *Journal on Optimization*, 8 (3):682–706, 1998.

Edward P Lazear. Culture and language. *Journal of Political Economy*, 107(S6): S95–S126, 1999.

Guillaume Lemaitre, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

Jonathan Levin. Wars of Attrition. October 2004. URL https://web.stanford.edu/~jdlevin/Econ%20286/Wars%20of%20Attrition.pdf.

Paul M. Lewis. *Ethnologue : languages of the world.* SIL International, 2009.

Sara Lowes, Nathan Nunn, James A. Robinson, and Jonathan Weigel. The evolution of culture and institutions: Evidence from the kuba kingdom. *Econometrica*, 2016.

Alexandre L Maltais. The tpp and cultural diversity. *The Trans-Pacific Partnership and Canada: A Citizen's Guide*, page 145, 2016.

Anna Maria Mayda and Dani Rodrik. Why are some people (and countries) more protectionist than others? *European Economic Review*, 49(6):1393–1430, 2005.

Jacques Melitz. Language and foreign trade. *European Economic Review*, 52(4): 667–699, 2008.

Stelios Michalopoulos. The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–39, 2012.

Stelios Michalopoulos and Elias Papaioannou. The long-run effects of the scramble for africa. *American Economic Review*, 106(7):1802–1848, 2016.

Stelios Michalopoulos, Louis Putterman, and David N. Weil. The influence of ancestral lifeways on individual economic outcomes in sub-saharan africa. *Journal of the European Economic Association*, 174(4):1186–1231, August 2019.

Chad Monfreda, Navin Ramankutty, and Jonathan A Foley. Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, 22(1), 2008.

David R. Mortensen, Patrick Littell, Akash Bharadwaj, Kartik Goyal, Chris Dyer, and Lori S. Levin. PanPhon: A resource for mapping IPA segments to articulatory feature vectors. In *COLING*, 2016.

Sendhil Mullainathan and Jann Spiess. Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

George Peter Murdock. *Africa : its peoples and their culture history*. McGraw-Hill, New York [u.a.], 1959.

J. Nocedal and S. Wright. *Numerical Optimization*. Springer New York, 2006.

Nathan Nunn and Nancy Qian. The potato's contribution to population and urbanization: Evidence from an historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650, 2011.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–3252, 2011.

Jacques Olivier, Mathias Thoenig, and Thierry Verdier. Globalization and the dynamics of cultural identity. *Journal of international Economics*, 76(2):356–370, 2008.

John C Paolillo. Language variation on internet relay chat: A social network approach. *Journal of sociolinguistics*, 5(2):180–213, 2001.

Robert D Putnam. E pluribus unum: Diversity and community in the twenty-first century. *Scandinavian political studies*, 30(2):137–174, 2007.

James E Rauch and Vitor Trindade. Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130, 2002.

Radim Rehurek and Petr Sojka. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50. ELRA, 2010.

Marta Reynal-Querol and Jose G Montalvo. Ethnic polarization, potential conflict and civil war. *American Economic Review*, 95(3):796–816, 2005.

David Ricardo. *On the Principles of Political Economy and Taxation*. John Murray, London, 1817. URL https://www.econlib.org/library/Ricardo/ricP.html.

Gillian Sankoff and Hélène Blondeau. Language change across the lifespan:/r/in montreal french. *Language*, pages 560–588, 2007.

Carol Schmid, Brigita Zepa, and Arta Snipe. Language policy and ethnic tensions in quebec and latvia. *International Journal of Comparative Sociology*, 45(3-4): 231–252, 2004.

Carol Myers Scotton and John Okeju. Neighbors and Lexical Borrowings. *Language*, 49(4):871–889, 1973.

Moses Shayo. Social identity and economic policy. *Annual Review of Economics*, 2020.

Laura Silver, Shannon Schumacer, and Mara Mordecai. In u.s. and uk, globalization leaves some feeling 'left behind' or 'swept up'. *Pew Research*, Octobeer 2020.

Enrico Spolaore and Romain Wacziarg. The political economy of heterogeneity and conflict. In Tibor Besedes and Volker Nitsch, editors, *Disrupted economic relationships: disasters, sanctions, dissolutions*, CESifo seminar series, pages 7–43. The MIT Press.

Enrico Spolaore and Romain Wacziarg. War and relatedness. *The Review of Economics and Statistics*, 98(5):925–939, 2015.

Morris Swadesh. Salish internal relationships. 16(4):157–167, 1950.

Guido Tabellini. The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950, 2008.

U.S. Official Inflation Data, Alioth Finance. Inflation Calculator. URL https://www.officialdata.org.

J.M. Vansina. *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa*. University of Wisconsin Press, 1990.

Hal R. Varian. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28, 2014.

Soren Wichmann, Eric W. Holman, and Cecil H. Brown. The ASJP database (version 17). 2016.

William Winkler. *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage*. 1990.

Appendix A. Hypothesis & Predictions Appendix (ONLINE APPENDIX)

A.1. Model Setup

We think about an individual earning some lifetime profit π if they do not adopt a new language, but they have the option of learning a new language at cost F in order to earn $\pi' \geq \pi$. The individual will adopt the new language if $\pi' - F \geq \pi$, and not otherwise.

We consider the case where a society i has a fixed set of neighbours ($j \in \mathcal{J} = \{1, \ldots, J\}$) which allows us to consider crowd-out among adopters of different languages.[64] Some of these neighbours represent viable trading partners, where both parties can gain from trade, and others are unlikely to economically interact. Individuals decide whether they want to become bilingual, with the following optimization problem:

$$max\{\pi, \pi'^1 - F \cdots, \pi'^J - F\}$$

The share of bilingual individuals in i (L_i) is weakly increasing in additional neighbours. We think of L_i as follows:

$$L_i = \sum_{j=1}^{J} L_{ij} \tag{9}$$

Where L_{ij} is the share of people in i for whom $max\{\pi, \pi^1 - F \cdots, \pi'^J - F\} = \pi'^j - F$.[65]

For a loanword word to diffuse, it needs to be used in a conversation, which requires that it must be both known by the using party and understood by the receiving party. Diffusion will therefore be increasing in use, and use will be an increasing function of probability of a conversation being between bilingual people. For any two people in the same society having a conversation, the probability that they share an adopted language j is: L_{ij}^2.

We assume that at rate ρ a conversation that could effectively use a loanword results in its diffusion. We now define: $\mathcal{L}_{ij} = \rho L_{ij}^2$ to denote the loanwords from j

[64]The main results of the model also hold in the much simpler two-language case, as in Supplementary Materials section SM-A.1.

[65]Here we are assuming that linguistic adoption has a negligible effect on these gains from trade. That is to say, no individual in i who would otherwise choose not to become bilingual would be incentivised to learn another language by the gains in communication with other bilinguals in their own language.

in a given language i.[66] We can rewrite everything as *shares* of bilingual citizens that adopt a particular language, so: $\ell_{ij} = \frac{L_{ij}}{L_i}$. Total loanword adoption is:

$$\mathcal{L}_i = \sum_{\mathcal{J}} \mathcal{L}_{ij} = \rho \sum_{\mathcal{J}} L_{ij}^2 = \rho L_i^2 \sum_{\mathcal{J}} \ell_{ij}^2 \tag{10}$$

Taking the derivative with respect to the share of speakers of a viable trading partner's language (called language j), and defining $\delta\ell_{iv}/\delta L_{ij} = -\sigma_{vj}$ (with $\sigma \in [0,1]$) as the rate at which language j crowds out the language of some $v \in \mathcal{J}$, we get:

$$\frac{d\mathcal{L}_i}{dL_{ij}} = \underbrace{2\rho L_i \left(\sum \ell_{iv}^2\right) \frac{dL_i}{dL_{ij}}}_{\text{Change in bilingualism}} - \underbrace{2\rho L_i^2 \sum \ell_{iv}\sigma_{vj}}_{\text{Change in composition of bilingualism}} \tag{11}$$

The first term represents the change to the share of bilingual speakers, and is weakly positive. Bilingualism will not decrease as a language provides more gains. The second term is the intensity of bilingualism in any given language. When intensity is high, diffusion is high.

A.2. Does more trade increase cultural convergence?

i) Trade-partner quality The derivative in equation 11 clearly has ambiguous sign, so the underlying conditions will matter for whether a society would be expected to converge or not with their trading partner (for details, see Supplementary Materials section SM-A.2).

First consider whether more gains from trade with language j (i.e. an improvement in partner quality) increases cultural adoption. Intuitively, there are two mechanism by which increased gains from trade with a given partner could impact total loanwords diffusion. First, by increasing individual adopters of that language, this increases total bilingualism which unambiguously leads to greater diffusion of loanwords. The second mechanism is by changing the intensity of bilingualism in a given language, and has an ambiguous effect, as increased adoption of a language may convert borrowers from other languages. In this case, if increased gains from trade leads to a language crowding out adopters of a popular language, this will lead to adoption being spread more thinly across many second

[66] The linguistics literature is heavily focused on determinants of diffusion, with concerns over potential for improvements in grammatical efficiency or filling vocabulary gaps, but for our purpose, since we are more interested in the adoption process (but do not want to ignore diffusion entirely) it is enough to abstract away from these considerations somewhat and keep diffusion very simple.

languages, slowing diffusion. We can only unambiguously say that this effect will increase loanword diffusion if we consider increased gains from trade with the most popular language. For any other language, it is possible that adoption will become converted away from high-intensity languages and instead become spread across more languages with lower intensity, so the second mechanism related to intensity of bilingualism may dominate the total bilingualism mechanism, and may decrease total loanword diffusion.

Prediction 1 (Quality matters)**.** *Total loanword diffusion is increasing with economic incentives for interaction with a society's best neighbour.*

ii) Trade-partner quantity Now we focus on the *quantity* of viable trading partners and assume all viable partners have the same *quality* and the same number of borrowers ($L_{ij} = \bar{L}_i \; \forall j \in \mathcal{J}$). Therefore:

$$\mathcal{L}_i = \rho \sum_{\mathcal{J}} L_{ij}^2 = \rho \sum_{\mathcal{J}} \bar{L}_i^2 = \rho J \bar{L}_i^2 \tag{12}$$

We denote the amount of crowd-out from each existing language to a new language as $\bar{\sigma}(J\bar{L}_i)$, which is increasing in the total number of bilingual individuals in a society. We take the derivative of diffusion $\mathcal{L}_i$ with respect to the number of partners J (see Supplementary Materials section SM-A.3):

$$\frac{d\mathcal{L}_i}{dJ} = \rho \left[\bar{L}_i^2 - 2J\bar{L}_i \bar{\sigma}(J\bar{L}_i) \right] \tag{13}$$

We show that at low levels of crowd-out, this is positive but as the number of bilinguals increases, saturation sets in and this derivative becomes negative (See Equations 32 and 33 in Supplementary Materials section SM-A.3). We then derive the second derivative of diffusion with respect to the quantity of neighbours ($\frac{d^2\mathcal{L}_i}{dJ^2}$) and show that it is negative (See Supplementary Materials section SM-A.4). So $\frac{d\mathcal{L}_i}{dJ}$ starts out greater than zero when crowd-out is low, which is the case when there are few bilinguals, but becomes negative as there are more bilinguals in the group and crowd-out increases. Taken together, these facts show that our framework predicts an inverse-U relationship between the number of viable trading relationships and total loanwords.

This analysis so far deals primarily with word adoption, but this framework can also describe the relationship between diversity and word *lending*. Intuitively, this substitution effect only matters for the *total* diffusion of loanwords and causes a nonlinear relationship where loanword borrowing is first increasing then decreasing in the number of viable trade partners. For the diffusion of loanwords from a given language, more adopters will unambiguously lead to more adoption of that given language.[67] Because of this, we do not get the same nonlinearity in lending as we do in loanword borrowing. Instead, loanword lending is nondecreasing in the number of adopters in other groups.

Prediction 2 (Quantity Matters)**.** *Total adoption is inverse-U shaped in the number of viable trading partners, while total lending is non-decreasing.*

iii) Conditions for convergence We use the following setup for the strategic learning decision as a standard war of attrition game played between two players from different neighbouring groups. Here, having the other partner bear the cost of language acquisition is analogous to 'winning' the war of attrition. Therefore the utility of winning for player j is $W_j = 0$, while the utility of losing for player j is $L_j = -F$ where F is the cost of learning language and avoiding this cost is the prize for winning the war.

The cost of staying in the war of attrition for player j is lost gains from trade, so $c_j = \gamma_j$ where $\gamma_j = \pi' - \pi$. Therefore the likelihood of winning a rational player must have per unit of time in order to be indifferent between staying in or exiting is $\lambda_j = c_j/p_j = \gamma_j/f$ where, assuming common F, the player who loses the most from continuing the war of attrition is the player with higher gains from trade γ, meaning they have the higher λ.

We now introduce the probability z_j that player j is a type that 'irrationally' will not exit, despite it being the rational decision. Therefore if the probability of exit if player j was fully rational is $\hat{G}(\cdot)$, then the actual probability of exit is $G(\cdot) = (1 - z_j)\hat{G}(\cdot)$.

If we consider groups i and j playing this game, where player j has the higher gains from trade, then $\lambda_j \geq \lambda_i$. For simplicity we also assume that $z_j = z_i = z$. As $\lambda_j > \lambda_i$, we also have that $\lambda_i \ln z_i \geq \lambda_j \ln z_j$.

Following Levin (2004), there is a unique perfect bayesian equilibrium for this war of attrition. In the case that $\lambda_i \ln z_i \geq \lambda_j \ln z_j$, this equilibrium is:

$$G_i(t) = 1 - e^{-\lambda_j t} \tag{14}$$

$$G_j(t) = 1 - z_j z_i^{-\lambda_i/\lambda_j} e^{-\lambda_j t} \tag{15}$$

where the 'weaker' player j exits immediately with positive probability. In our

[67] See Supplementary Materials section SM-A.5 for a formal statement of this result

situation, this means that j likely loses the war of attrition and bears the cost of learning the language of group i:

Prediction 3 (Asymmetric Incentives Matter)**.** *For a pair of neighbours, the group that gains the most from trade will be the society that bears the cost of acquiring common language.*

A.3. Contact hypothesis

We compare our framework to the *contact hypothesis*, an alternate theory of cultural change that has been discussed broadly in the literature. One main difference between the two is that in our model individuals purposefully reduce their cultural distance to induce or facilitate trade. Under the contact hypothesis, first societies interact, and then as an unintended by-product, they adopt elements of each other's cultures. The predictions under each are different, so for policy-makers to understand when and to what extent to take concerns of cultural threat seriously, it is important to understand which model better fits the data.

The contact hypothesis suggests that any trade will involve both adoption and lending, unlike our framework which predicts strong asymmetries in adoption. However, while adoption is symmetric under the contact hypothesis, diffusion may not be. There could be asymmetries in loanwords despite symmetric adoption because, for instance three adopters in a society of four will generate more diffusion than in a society of one-hundred. Asymmetry in loanwords in the contact hypothesis therefore depends on relative population size rather than trade leverage.

We now formulate this alternative hypothesis in terms of the the framework outlined in Section 3, and generate two alternative predictions we will refer to in the discussion of our empirical results.

Diffusion of loanwords under the hypothesis occurs when someone in group i interacts with someone else in group i who either learned j themselves, or has a partner in group j who made the adoption investment to initiate trade. Therefore, the number of people in group i who may diffuse loanwords is the total number of i-j partnerships regardless of who made the linguistic investment, giving the following diffusion function:

$$\mathcal{L}_{ij} = \rho \left(\frac{N_i L_{ij} + N_j L_{ji}}{N_i} \right)^2 \tag{16}$$

which gives us that

$$\frac{d\mathcal{L}_{ij}}{dL_{ij}} = 2\rho \left(\frac{N_i L_{ij} + N_j L_{ij}}{N_i} \right) > 0 \tag{17}$$

and so, under the contact hypothesis, loanword diffusion into language i is increasing in the number of individuals in language j who adopt language i.

Prediction-CH 1 (Contact Hypothesis: Symmetric Incentive Significance)**.** *Under the contact hypothesis, holding group i's expected gain from trade fixed, an increase in group j's expected gain from trade will also lead to an increase in loanwords from language j into language i.*

In addition, if we consider a given pair of languages, we have that

$$\mathcal{L}_{ij} > \mathcal{L}_{ji} \iff N_j > N_i \tag{18}$$

Prediction-CH 2 (Contact Hypothesis: Size-Driven Asymmetry)**.** *Under the contact hypothesis, the only source of asymmetry in adoption rates within a pair of neighbours is different population size. Therefore, controlling for the size of groups, there should be no asymmetry in linguistic adoption within a given pair.*

Appendix B. Lexical Data & Loanword Prediction (ONLINE APPENDIX)

B.1. Overview

The preparation of the dataset follows the following rough order. We first extract our data and *epitranscribe* the orthographic representation into phonetics. We then compute own-language dissimilarity measures to identify words that look like outliers. Then we match to data on contextual similarity and use this to construct candidate word-pairs that are in the same semantic space for which we then compute additional pairwise distance measures between words. We then use all these features computed on our training set to train our classifier algorithm which we then apply to the full PanLex database. In figure B1 we graphically outline the workflow and how different datasets are used.

B.2. Data Extraction and Phonetic Transcription

The first task in creating this dataset was extracting data on expressions from the PanLex dataset, after which we prepared the necessary features for each expression, and transcribed orthographic representations into phonetic representations.[68] Some language families were not represented. We therefore coded orthographic-phonetic mappings using orthography tables from OmniGlot for 15 further languages, to give full coverage of the major language families included in our sample. We then use Ethnologue data on language families to match all languages in our sample to the nearest language sharing the same script included in our augmented list.

For each language, we build a dataframe including all expressions and extract the following information for each expression: *Unique Expression ID*, *Raw Text*, *Degraded text (no accents, etc.)*, *Language code*, and *Epitranscribed raw text.* Meaning identifiers in the PanLex dataset refer to abstract meanings, that may be associated with one or more expressions. If two expressions are assigned the same meaning identifier, they can be thought of as translations.

B.3. Train Machine Learning Classifier

For each word pair, we calculate the features described below, which are the inputs to the machine learning algorithm. These are the features the classification algorithm uses to decide if a given word-pair is a loanword transfer.

i) Own-Dissimilarity Measures A core requirement for identifying loanwords is that we can determine which words appear to be outliers in their language, that are likely to have been adopted, and which ones are unlikely to have been introduced from another language. We therefore generate the following measures of own-language dissimilarity.

The Jaro-Winkler metric computes the minimum edit distance between two words, accounting for transpositions, where greater weight is given to characters near the beginning of the word (Jaro, 1989; Winkler, 1990). As loanwords are likely to be adapted with added suffixes, this metric is suitable for measuring likelihood of a word being introduced from another language. This measure is between 0 and 1, with 1 being identical spellings. Our measure computes the Jaro-Winkler distance with other words in the same language, and we use the deciles of this distribution as features. We also construct this restricting only to the spellings of words in the language with similar meanings, using the same

[68]We used the `Epitran` package[69] to convert orthographic text into International Phonetic Alphabet (IPA), which relies on mappings between orthographic and phonetic units. `Epitran` includes 64 language-script pairs. For example 'eng-Latn', for English in Latin script, and 'tir-Ethi' for Tigrinya in Ethiopic script

threshold for contextual similarity as when we generate word-pairs, and compute quintiles of this distribution as features. For these contextually-similar words we also compute the phonetic difference and also use quintiles of this distribution as features.

We construct measures of whether the combinations of phonetic units, or *phonemes*, that make up a word are typical for the language. Using the phonetic transcriptions of PanLex expressions, we build a list of all 2- and 3-grams of phonemes contained in a language and compute the expected number of occurrences of this n-gram in words, and the position of this n-gram in words, from that language. For each word, we then take the average of this score for all contiguous sequences of two or three phonemes making up a word.

In the basic phonetic n-gram measure we create above, we create an expected occurrence score for 2- and 3-grams of a word based on observed occurrence in all words in the language. To improve this measure, and compare words to the 'core' words in a language that are highly unlikely themselves to be loanwords, we construct a similar expected occurrence score for 2- and 3-grams based on observed occurrence in words that are part of the Swadesh list for that language.

We therefore construct measures of whether the combinations of phonetic units, or *phonemes*, that make up a word are typical for words from the Swadesh list for a language. Our source of Swadesh words is the 40 word lists compiled as part of the Automatic Similarity Judgement Program (Wichmann et al., 2016). Using the phonetic transcriptions of these Swadesh words, we build a list of all 2- and 3-grams of phonemes contained in a language and compute the expected number of occurrences of this n-gram in Swadesh words, then take the average of this score for all contiguous sequences of two or three phonemes making up a word.

To restrict the space of candidate word pairs we consider, we generate a measure of the contextual distance between concepts. To do so, we use a pre-trained model of word vectors trained from the Google News dataset. This model has a vocabulary of roughly three million expressions, and can be used to generate measures of contextual similarity for English words.[70] For all meanings in the PanLex dataset with an English denotation, or a definition in English, we can assign a contextual similarity score, between 0 and 1. For all expressions with the same meaning identifier, we assign a similarity score of 1.

This word vector measure of contextual similarity of expressions is less restrictive than considering only to expressions that are translations, and broadens the space of potential loanword pairs while preventing nonsensical matches between

[70] This contextual similarity is implemented by the `Gensim` package (Rehurek and Sojka, 2010).

expressions denoting entirely unrelated concepts.

ii) Word-Pair Construction and Pairwise Distance Having created own-language dissimilarity measures for expressions and mapped them into space of contextual similarity, we create word-pairs that are candidates for being loanwords, and generate pairwise distance measures. We restrict our analysis to expressions who are above a threshold of contextual similarity, set at 0.7. This threshold is low enough that we consider a broad range of related meanings, but is high enough to be practical and reduce the number of comparisons made to a level that can be carried out with a reasonable amount of computing time.

For each word in our dataset, we create pairwise matches with all words in all other languages. As each *expression* may be mapped to multiple *meanings*, we create pairwise matches at the word-meaning level, and restrict to the most similar meaning pair for each word-pair where words have multiple meanings. We then restrict to pairs of words that are contextually similar, as above. We then calculate a number of pairwise distance measures between the two words, as follows.

iii) Articulatory Feature-Edit Distance Metrics The first set of pairwise distance metrics we create is exploits detailed information on the phonemes that make up the phonetic representations of words. We map each phoneme to a vector of twenty-one articulatory features describing the way a spoken sound is actually produced, such as tongue position, open or closed mouth, etc.[71] This level of detail means that phoneme differences can be weighted by how similar the two phonemes sound. Using these articulatory vector representations, we construct two pairwise minimum edit distances. The Hamming Feature-Edit Distance computes the minimum distance between two words, allowing for insertion and deletion of phonemes and accounting for the difference in phonemes weighted by difference in articulatory features. The Weighted Hamming Feature-Edit distance is similar to the unweighted Hamming distance, but where the cost of articulatory feature edits are differently weighted depending on their class and subjective variability.

iv) Jaro-Winkler Distance As with the own-language dissimilarity measures, we compute the Jaro-Winkler orthographic distance for the candidate wordpair.

v) Language Family Cladistic Distance For the candidate wordpair, we also compute the pairwise cladistic distance between the two languages. This data is based

[71]This is done using the PanPhon package developed in Mortensen et al. (2016)

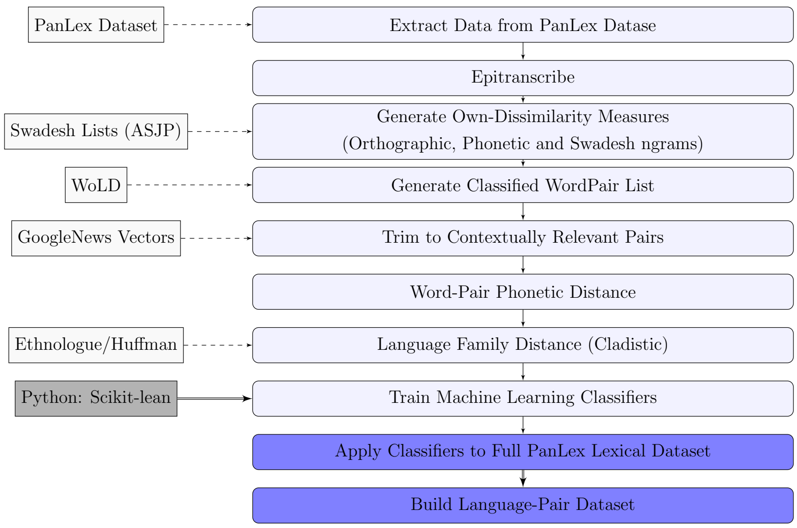
on the Ethnologue language family trees Lewis (2009), where the measure of linguistic family distance is equal to the share of nodes in the first language's tree that are also in the second language's family classification.

vi) Pairwise Difference in Own-Language Dissimilarity In addition to these measures of pairwise difference between words, we also calculate the *difference* in all of the own-language dissimilarity measures generated above. By including the differences in these measures as features in the machine learning algorithm, we allow the classifier to explicitly decide whether one word in a pair appears more likely to be an outlier than the other.

B.4. Classified Training Sets

Having created these features, we match the the World Loanwords Database (WoLD) of words with manually classified origins to the dataset of PanLex words.

Figure B1: Code & Data Flowchart



Appendix C. Trade Model

Societies choose an allocation of land ($\vec{l}$) to different crops, and output is land allocated to a crop multiplied by productivity, where the productivity vector ($\vec{q}$) is the average from the GAEZ dataset described in Section 4.B.

$$Y(\vec{q},\vec{l}) = [y_0(q_0, l_0), \cdots, y_{41}(q_{41}, l_{41})] = [q_0 \cdot l_0, \cdots, q_{41} \cdot l_{41}] \tag{19}$$

We define a nutritional utility function that takes a Cobb-Douglas form.

$$U(x_0, x_1, \cdots, x_{16}) = x_0^{\alpha_0} x_1^{\alpha_1} \cdots x_N^{\alpha_{16}} \tag{20}$$

where x_0 represents daily calories, and x_1 through x_{16} are the sixteen essential micronutrients. The weights for essential nutrients, α_i, are constructed as follows: $\alpha_i = \frac{\gamma_i}{\sum_j \gamma_j}$ where we use the Daily Reference Intake (DRI) amounts as γ_i, for $i \in \{1, 2, \cdots, 16\}$ and $j \in \{0, 1, 2, \cdots, 16\}$, where we normalize the weights so that the exponents sum to one. and so they capture the relative importance of nutrients in the diet.

For γ_0, the weight for calories, we calibrate using observed population figures. This is because the DRI figures we use are derived from modern North American diets, and it is not reasonable to assume that the implied tradeoff in macro- and micro-nutrients can be generalized to the preindustrial local trading systems we are trying to approximate.[72] Calibrating γ_0 in this way is intuitively similar to capturing the quality-quantity tradeoff between satisfying all nutrient requirements in modern diets, and having a greater population.

i) Trade We first numerically solve for full-trade production, assuming full efficiency of each localized trading group. This produces a forty-one dimensional vector $\vec{l} = [l_0, l_1, \cdots, l_{41}]$ of land share l_c allocated to crop c that maximizes the nutritional utility function in Equation 38 for all groups in a neighbourhood, subject to a constraint on each group's land allocation shares.[73]

We then solve for the set of equilibrium prices supporting these land allocations under trade, and compute the budget of each group in the neighbourhood. See Supplementary Materials section SM-D.4.1 for details on this process. Given the properties of our utility function above, all groups consume in the same proportions, so their individual consumption will be their share of the neighbourhood's total budget times the total crop output of the neighbourhood. From this consumption bundle, we then compute utility under trade.

Appendix D. Word-Topic Identification

In order to identify words that are associated with topics of particular interest across the huge range of languages included in our dataset, we start from word

[72]For a discussion of model validation, please se Section SM-D.5.

[73]see Supplementary Materials section SM-D.4 for details on data cleaning that were undertaken prior to estimation and on the exact estimation algorithm.

lists in English and use translations as well as word-vector models to identify contextually similar words in as many languages as possible.

First, we chose lists of words that represented the topics of interest. These topics are as follows: transactions, technology, religion, politics, gender & broader human rights, social organization, as well as crop names. For crop names, we simply used the names of the crops included in the FAO GAEZ dataset. For other topics we chose lists of ten words, listed in Table D1 below, that were as specific to the topic as possible (with the fewest misleading multiple connotations) as well as being broadly applicable across cultural contexts (i.e, not choosing words only relevant to a few cultures). We then matched these wordlists to the PanLex dataset for English. The PanLex dataset is organized such that each expression (where an expression is a *lexical* object, the unique combination of letters) is matched to multiple meaning codes.[74] We match by spelling from our chosen wordlists to expressions in PanLex, and keep their associated meaning identifiers.

To expand beyond English, we use these meaning identifiers to find direct translations in other languages. We first identify direct matches by meaning ID into the 293 other languages covered by contextual similarity models by identifying all expressions with exactly the meaning IDs identified in English. We then add all these transitively matched meaning-IDs associated with these expressions to the list of meanings matched to our original topic word. At this point we now have a combined list of meanings directly matched to each original topic word, and those matched transitively in the 293 other languages. To go beyond direct translations, we take the combined list of meaning identifiers described above, and for each of the languages, we extract a list of expressions (recall that these expressions are unique *spellings* or *lexical* objects) corresponding to these meanings. We then run these expressions through the contextual similarity model for each language[75] to identify words that are not direct translations, but are very contextually similar at different thresholds. For crop names, we set a high threshold, as we are translating specific words, but for the other topic groups we use a lower threshold as we are capturing a broader range of concepts. Having identified contextually similar *expressions*, we then match these back to the meaning identifiers in PanLex for each language. Combining these lists of meaning identifiers then gives us the final,

[74] For example, the expression '*bolt*' is matched to meaning identifiers denoting the 'hardware' meaning, as well as the 'sprinting' meaning.

[75] These models are generated using fastText word-vectors on Wikipedia pages for each language and for a pair of expressions, will determine the contextual similarity of the two words. See Bojanowski et al. (2017) for details on the methodology. Word vectors retrieved from https://fasttext.cc/docs/en/pretrained-vectors.html on May 23, 2020.

full list of meaning identifiers that represent the topic of interest. We use these meaning identifiers to categorize loanwords and generate a count of how many words we identify in a language as being related to a particular topic. This allows us to empirically examine the share of words in a language tagged as related to a given topic that are adopted, and who they are adopted from.

Table D1: Topic Categories and Origin Wordlists

Topic	**Starting Word List**
Transactions	money, expensive, price, trade, exchange, loan, delivery, buy, product, contract
Technology	plough, book, boat, harvest, irrigation, medicine, map, machine, planting, fishing, husbandry
Religion	God, priest, afterlife, spirit, pray, worship, sacred, church, temple, mosque, astrology
Politics	leader, ruler, capital, government, policy, law, council, jurisdiction, justice, authority
Gender & Human Rights	queen, housewife, equality, slave, freedom, agency, empowerment, chivalry, child labour, effeminate
Social Organization	polygamy, polygyny, marriage, husband, wife, adoption, cousin, inheritance, ancestor, ancestry, kinship
Crop Names	alfalfa, banana, plantain, barley, buckwheat, cacao, canary grass, carrot, cassava, manioc, chickpea, lemon, lime, orange, coconut, cotton, cowpea, dry pea, flax, foxtail millet, millet, green grams, groundnut, peanut, maize, corn, miscanthus, silvergrass, oat, palm tree, oil palm, palm, olive, onion, phaseolus, bean, pigeon pea, pea, rye, sorghum, soybean, soya, beet, sugarbeet, sugarcane, sugar, sunflower, sweet potato, sweetpotato, switchgrass, bunchgrass, tea, tobacco, tomato, rice, wheat, potato, yam

Note: This table shows the topic categories we look at, and the starting lists of English words used to propagate meanings through our broad range of languages.

Appendix E. Supporting Evidence (ONLINE APPENDIX)

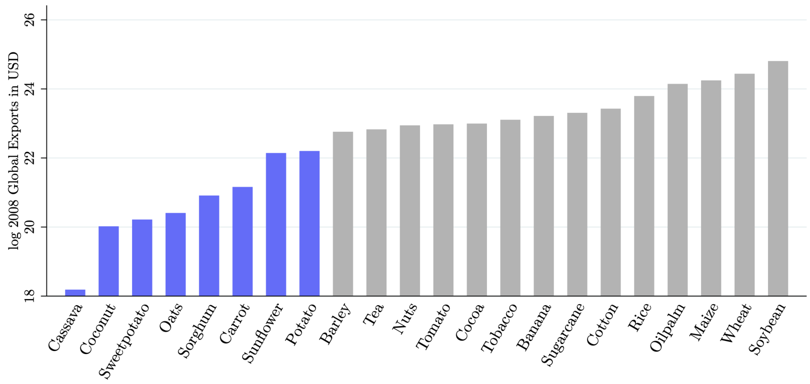
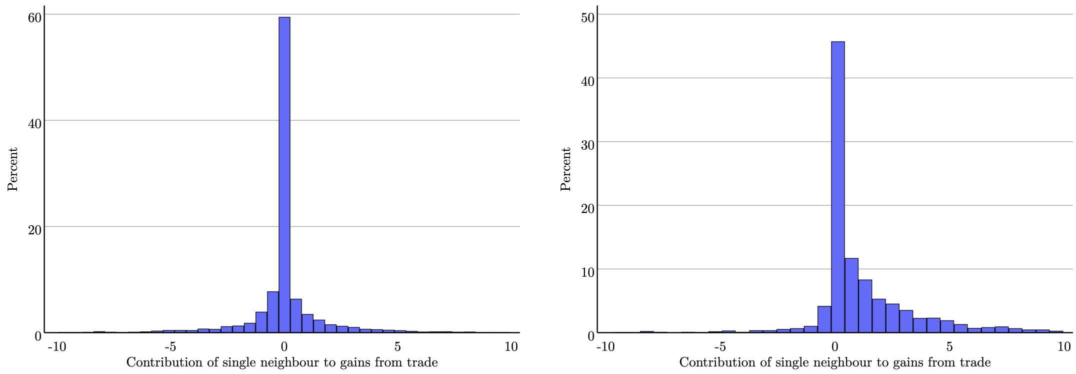


Figure E1: Delineation of global crops used for trade model validation

Note: The figure shows the log exports of 2008 trade as per the FAO. We use these trade numbers to determine which crops are globally traded and therefore not relevant to our regional trade model. All crops coloured grey are determined to be global crops, and we chose this cutoff based on the small break in the data on either side of the cutoff.

Source: Author constructed based on data from FAO accessed August 12, 2020. http://www.fao.org/faostat/en/#data/TP



(a) Relationship-level **(b)** Society-level

Figure E2: Histogram of Gains From Trade

Note: The figure shows histograms of the output of the trade model. On the top we have the bilateral measure of gains from trade and on the bottom we have the societal level measure. Since on the top we compare full-trade with a partner to full trade without a partner, the measure can be greater or less than 1, and in fact it seems centred around 1 - indicating that most societies are indifferent towards the inclusion of most of their neighbours in their trading network. A value of less than one indicates that the society is worse-off due to the existence of their neighbour in their trading network - i.e. the societies are economic competitors. A value of greater than one indicates that we expect the societies have a profitable trading relationship. On the bottom we show our societal level measure which compares trade with the network as a whole to autarky. Here we only see values greater than one because societies always have the option of behaving as if there exists no trading network. By far most societies have values less than two. A value of two indicates that the society is twice as well off with the existence of their trading network relative to autarky.
Source: Author constructed. Data sources are described in the text.

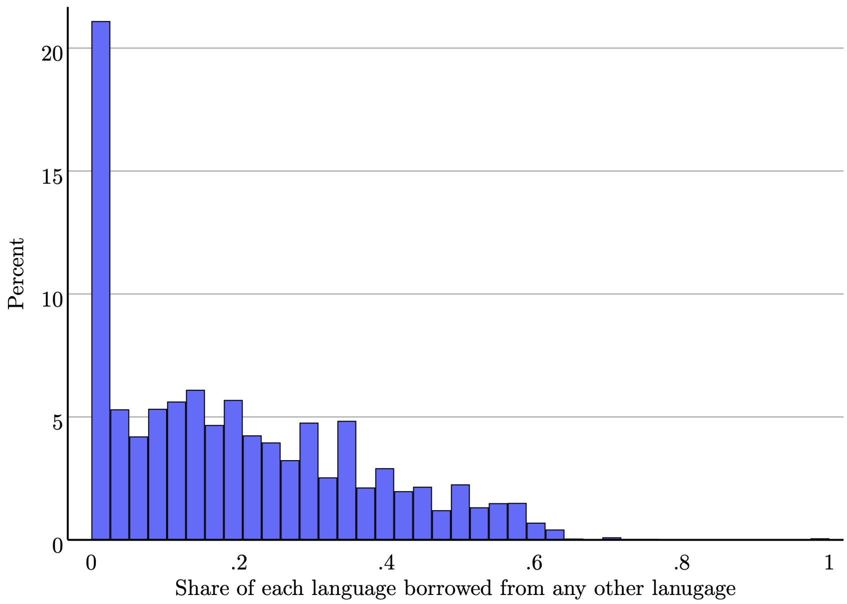


Figure E3: Histogram of Language Borrowing

Note: The figure shows the raw-data of the main dependent variable used throughout the paper, the share of any given language adopted from one of their neighbours. Notably, while about 20% of societies do not adopt at all, a non-trivial share of societies adopted between 20% and 60% of their language. This justifies a focus on loanwords, and illustrates that it is a non-trivial source of variation in linguistic distance.
Source: Author constructed. Data sources are described in the text.

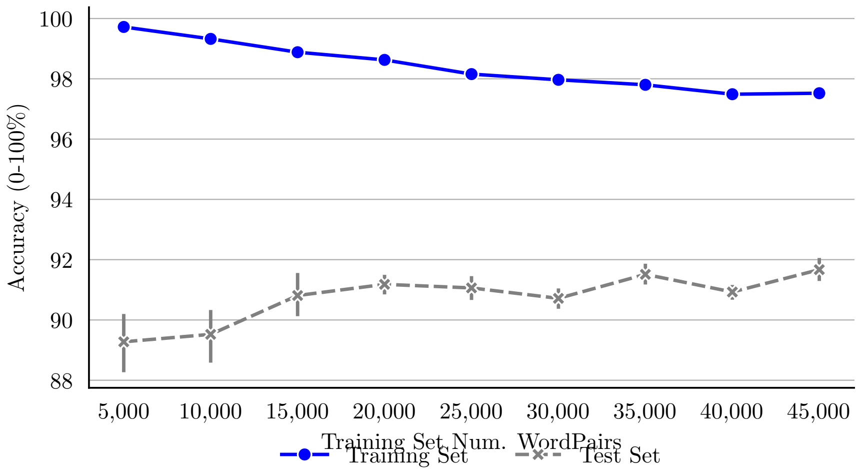


Figure E4: Accuracy of Phase Two Classifier

Note: This graph shows the results of the second-stage classifier described in the text. The second stage classifier refines the predictions made by the first-stage classifier by focusing on word-pairs the first stage classifier identifies as loanwords and improving the accuracy of our predicted loanwords.

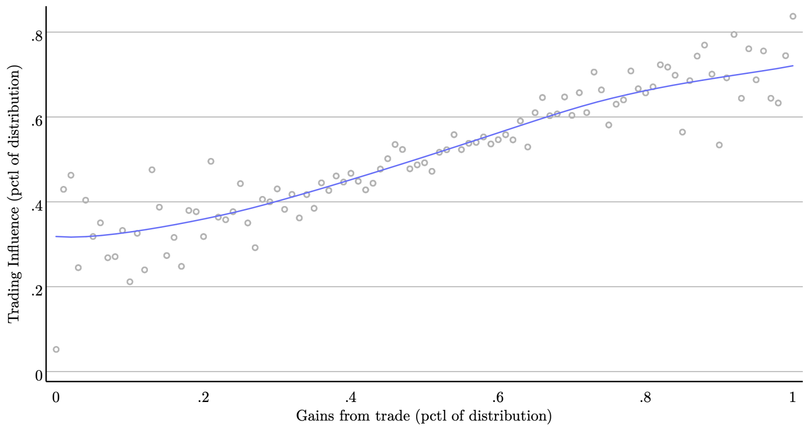


Figure E5: Gains from trade and trade influence

Note: The figure shows the correlation between gains from trade and trade influence. The graph uses a society as the unit of observation, and takes the mean across neighbours for both gains and influence. The scatterplot groups observations into 0.01 gains from trade bins. The fit line in each graph is based on a biweight kernal of degree 1, with a bandwidth of 0.025.

Table E1: Robustness: loanword thresholds of 60% & 70% instead of 50%

Dependent Variable:	Language Adoption		Language Influence	
Threshold Used	60%	70%	60%	70%
	(1)	(2)	(3)	(4)
Gains from trade with neighbours	0.00361**	0.00156*		
	(0.00169)	(0.000801)		
Influence on trade with neighbours			0.0114***	0.00382**
			(0.00403)	(0.00175)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
N	2,597	2,597	2,597	2,597
R^2	0.069	0.112	0.112	0.122
Dependent Variable Mean	0.00426	0.00144	0.00423	0.00142

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. To differentiate by pre/post Colombian exchange, we recalculate all gains from trade computations, restricting the analysis only to the pre-Colombian exchange crops. To get the post-Colombian exchange measures we subtract our unconstrained measure from the pre-Colombian exchange measure. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours.

Table E2: Gains from trade with best neighbour and language exchange with other neighbours

	Exchange with best		Average exchange		Exchange with worst	
	Borrowed	Loaned	Borrowed	Loaned	Borrowed	Loaned
	(1)	(2)	(3)	(4)	(5)	(6)
Gains from trade with neighbours	0.0603**		0.00942		-0.00561	
	(0.0299)		(0.0131)		(0.00994)	
Influence on trade with neighbours		0.162***		0.0197*		-0.00967*
		(0.0599)		(0.0117)		(0.00563)
Trade wealth (structurally estimated)	✓	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓	✓
Land Share	✓	✓	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓	✓
N	2,602	2,602	2,602	2,602	2,602	2,602
R^2	0.062	0.133	0.038	0.042	0.015	0.013
Dependent Variable Mean	0.800	0.635	0.292	0.210	0.0791	0.0492

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue.

Table E3: Colombian Exchange: Are results driven by long or shorter run trade?

Dependent Variable:	Language Borrowed			
	(1)	(2)	(3)	(4)
Gains from trade - post-Colombian exchange crops	0.652***		0.701***	
	(0.182)		(0.225)	
Gains from trade - pre-Colombian exchange crops		0.611***		0.473**
		(0.175)		(0.198)
Trade influence - post-Colombian exchange crops			-0.116	
			(0.216)	
Trade influence - pre-Colombian exchange crops				0.274
				(0.223)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
N	2,600	2,600	2,600	2,600
R^2	0.062	0.061	0.062	0.061
Dependent Variable Mean	0.799	0.799	0.799	0.799

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours.

Table E4: Falsification: Unviable Trading Relationships

	Language adoption			
Utility measure	percent change (1)	percentile rank (2)	percent change (3)	percentile rank (4)
Gains from trade with neighbours	-0.0333	-0.429	-0.0330	-0.442
	(0.0388)	(0.266	(0.0389)	(0.289)
Influence on trade with neighbours			0.0242	-0.0543
			(0.0314)	(0.214)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
Observations	1,783	1,783	1,783	1,783
R-squared	0.134	0.135	0.148	0.150
Dependent Variable Mean	0.503	0.503	0.503	0.503

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of from the society's neighbours.

Table E5: Borrowing & viable partners not controlling for total partners

Dependent Variable:	Language Borrowed		Language Loaned	
	(1)	(2)	(3)	(4)
Number of viable trading neighbours	0.117***	0.317***	0.546**	0.292
	(0.0385)	(0.0673)	(0.214)	(0.323)
Number of viable trading neighbours squared		-0.0183***		0.0233
		(0.00510)		(0.0459)
Trade wealth (structurally estimated)	✓	✓	✓	✓
Population	✓	✓	✓	✓
Land Share	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓
N	2,602	2,602	2,602	2,602
R^2	0.067	0.072	0.150	0.154
Dependent Variable Mean	0.938	0.938	0.938	0.938

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We lose 229 observations relative to the sample in table 4 because there are some languages that were not colonized, and have a missing colonial centrality value. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5. Similarly, Influence on trade with neighbours is the analogous measure for lending. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours.

Table E6: Robustness to different thresholds

Dependent Variable:	Language Borrowed			
	60% threshold		70% threshold	
	(1)	(2)	(3)	(4)
Gains from trade with neighbours	0.00639**	0.00353**	0.00286**	0.00105
	(0.00303)	(0.00176)	(0.00112)	(0.000789)
Influence on trade with neighbours		0.00397		0.000624
		(0.00270)		(0.000967)
Relationship Fixed Effects	✓		✓	
Society Fixed Effects (both)		✓		✓
N	5,588	5,588	5,588	5,588
R^2	0.502	0.582	0.504	0.716
Dependent Variable Mean	0.00126	0.00126	0.000443	0.000443

Note: The unit of observation is a society-pair. Standard errors two-way clustered by each society within a society-pair are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_{ij} as defined in equation 4. Influence on trade with neighbours is the percentile rank in the distribution (range [0,1]) of ι_{ij}. Viable trading relationships are any relationships where at least one of the two parties can gain from trade. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. 'Society Fixed Effects (both)' means we separately include a society-fixed effect for each society in the relationship; 'Relationship Fixed Effects' means we include a fixed effect for a specific pair.

Table E7: Colombian Exchange: Are results driven by long or shorter run trade?

Dependent Variable: Sample:	Language Borrowed Viable Only	
	(1)	(2)
Gains from trade - post-Colombian exchange	0.226	
	(0.222)	
Gains from trade - pre-Colombian exchange		0.167
		(0.173)
Relationship Fixed Effects	✓	✓
N	5,543	5,543
R^2	0.520	0.520
Dependent Variable Mean	0.278	0.278

Note: The unit of observation is a society-pair. Standard errors two-way clustered by each society within a society-pair are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_{ij} as defined in equation 4. Similarly, Influence on trade with neighbours is the percentile rank in the distribution (range [0,1]) of ι_{ij}. Viable trading relationships are any relationships where at least one of the two parties can gain from trade. To differentiate by pre/post Colombian exchange, we recalculate all gains from trade computations, restricting the analysis only to the pre-Colombian exchange crops. To get the post-Colombian exchange measures we subtract our unconstrained measure from the pre-Colombian exchange measure. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. 'Society Fixed Effects (both)' means we separately include a society-fixed effect for each society in the relationship; 'Relationship Fixed Effects' means we include a fixed effect for a specific pair.

Table E8: Loanwords and trade incentives among unviable relationships

Dependent Variable: Sample:	Language Borrowed Unviable Only		
Utility measure:	percent change	percentile rank	
	(1)	(2)	(3)
Gains from trade with neighbours	-0.0120	0.205	0.193
	(0.0608)	(0.297)	(0.303)
Influence on trade with neighbours			-0.141
			(0.510)
Society Fixed Effects (both)	✓	✓	✓
N	3,830	3,830	3,830
R^2	0.727	0.727	0.727
Dependent Variable Mean	0.234	0.234	0.234

Note: The unit of observation is a society-pair. Standard errors two-way clustered by each society within a society-pair are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_{ij} as defined in equation 4. Similarly, Influence on trade with neighbours is the percentile rank in the distribution (range [0,1]) of ι_{ij}. Unviable trading relationships are any relationships where neither of the two parties can gain from trade. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. 'Society Fixed Effects (both)' means we separately include a society-fixed effect for each society in the relationship; 'Relationship Fixed Effects' means we include a fixed effect for a specific pair.

Table E9: Loanwords by word-type and trade incentives

	(1)	(2)	(3)	(4)	(5)
	Borrowing of word-type:				
	Politics	Religion	Human rights	Social org.	Technology
Gains from trade with neighbours	0.0407**	0.124***	0.0394***	0.0612	0.0770***
	(0.0185)	(0.0364)	(0.0124)	(0.0376)	(0.0253)
Trade wealth (structurally estimated)	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓
Land Share	✓	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓	✓
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓	✓
Colonizer FE	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓
N	2,504	2,504	2,504	2,504	2,504
R^2	0.014	0.025	0.014	0.020	0.018
Dependent Variable Mean	0.0446	0.128	0.0387	0.109	0.0815

Note: The unit of observation is a society. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We lose 85 observations relative to the sample in table 4 because there are some languages where we find no english equivalents for one category of word-type, and these observations are dropped across all specifications to facilitate comparisons within the table. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, which is aggregated to the society level by taking the maximum value from the society's neighbours. All word-type borrowing outcomes are winsorized at the 0.1% level to deal with outliers.

Appendix F. Supplementary Empirical Exercises (ONLINE APPENDIX)

F.1. The importance of colonialism for language exchange

It seems incomplete to think about language acquisition by societies without considering the obviously huge impact of colonial contact. This section offers some suggestive evidence of the influence of (a particular type of) colonial interaction on the acquisition of languages. In order to assess the consequences of colonial interaction, we need exogenous variation. One possibility that we pursue, is that a colonizer may strategically focus their presence closer to the centre of contiguous, populated blocks under their control. This may be strategically advantageous as a cost minimization strategy, to most efficiently provide services to (or alternatively exploit) the regions under their control. The centre of these regions may be exogenously determined, since the borders of the regions themselves are known to have been near-random (Michalopoulos and Papaioannou, 2016).

We start by identifying centroids for each functionally contiguous cluster using data on colonial history from Hensel and Mitchell (2007). We restrict to populated regions by only considering areas with an estimated potential caloric yield above 1000kcal (in the spirit of Galor and Özak (2015)).[76] Based on these regions, we identify cluster centroids, and construct the distance from the centroid of each society to the contiguous colonial centroid (see figure SM-F3). We expect the

[76] We further split clusters connected by narrow 'bridges' using small buffer zones to avoid overlap.

more 'central' a society, the more interaction they may have with a colonist, and the more of the colonial language they may adopt.

From our perspective, the key is not necessarily to understand whether colonial contact influenced colonial linguistic adoption (though that is independently of interest) but rather to understand whether colonialism could be an alternate mechanism for our core results. For that to be the case, it would need to be that gains from local agricultural trade was influencing regional language adoption through colonialism. In other words, it would need to be true that colonial intensity happened to be correlated with both gains from local agricultural trade, and with regional language adoption.

This may not be as far fetched as it first appears. It seems reasonable to consider that colonists may have decided to interact most with groups that were already regionally influential. In our case this would imply that colonial intensity might be higher for those with greater economic or cultural influence: those with greater trade or linguistic influence.

Table F1: Colonialism, local trade and regional language exchange

Dependent Variable:	Colonial language adoption (1)	Gains from trade (2)	Trade influence (3)	Language Borrowed (4)	Language Loaned (5)
Colonial centrality	0.0405***				
	(0.0136)				
Colonial language adoption		0.0161	0.00176	0.0239	0.0368
		(0.0107)	(0.00905)	(0.0200)	(0.0234)
Trade wealth (structurally estimated)	✓	✓	✓	✓	✓
Population	✓	✓	✓	✓	✓
Land Share	✓	✓	✓	✓	✓
Land diversity	✓	✓	✓	✓	✓
Gains from trade with neighbours	✓				
Distance to Neighbour(s) (quintic polynomial)	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓
N	2,211	2,235	2,235	2,235	2,235
R^2	0.025	0.013	0.018	0.032	0.370
Dependent Variable Mean	1.934	0.984	0.816	0.860	0.825

Note: The unit of observation is a society. Robust standard errors in parentheses, clustered by colonizer. *** p<0.01, ** p<0.05, * p<0.1. We lose 229 observations relative to the sample in table 4 because there are some languages that were not colonized, and have a missing colonial centrality value. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Distance to neighbours is a mean distance to neighbours, and in this case captures the density of the neighbourhood. Language Borrowed (range [0,100]) is defined in equation 1, while Language Loaned (range [0,100]) is the lending analogue. In each case we aggregate to the society level by taking the sum of the society's neighbours. For the p-value we compare standardized z-scores to account for the possibility that the lending and borrowing distributions are different. The results are nearly identical using the unstandardized values.

First, we check to see if colonial centrality influences colonial language adoption. In table F1 column 1 we regress colonial language borrowing on colonial centrality, as described above, with all of the controls from our main specification. We find, consistent with our hypothesis that colonial centrality does strongly positively influence colonial intensity. However, figure SM-F4 suggests that there is not a concern for our main results, since colonial centrality is not related to regional trade gains or trade influence. One caveat is this is based only on the *exogenous* variation in colonial presence while our concern is that colonists *endogenously* interacted more with economically influential groups. Accordingly, in columns 2-5 of table F1 we look to see if there is a correlation between colonial language borrowing and gains from trade, trade influence, regional language borrowing or lending. Again though, there is little evidence of any relationship.

F.2. Do the same factors influence diversity?

One initial motivation of the project was an investigation into whether and how diversity might be endogenous. This has been a very difficult issue for the literature, for two main reasons. First, data on investments in diversity have never existed, and second, exogenous variation has been difficult to find. We make progress on the second by structurally estimating, for each society pair, the gains from having (or not) that society in their agricultural trading network. This is straightforward, especially for our very simple trade model, but is computationally intensive.[77] The first issue we resolved by looking at loanwords as a potentially reasonable proxy for reductions in diversity. However, we now would like to ensure that loanwords respond in the same way as the broader measures of diversity used throughout the literature to our exogenous variation.

We aggregate our trade data to the country level. We do this by taking the predicted level of language exchange based on the model estimated in column 1, table F1 (based on equation 6) and taking the within-country mean of that predicted value. We then match that predicted value to the replication data from Alesina et al. (2016).[78] This allows us to estimate the exact specification from one of the most recent contributions to the literature.

[77] This issue was resolved with access to Compute Canada's Niagara supercomputer at the University of Toronto, one of the largest supercomputers in the world.

[78] We downloaded this data from the Journal of Political Economy website.

Table F2: Country-level: Endogenous diversity

	ELF (1)	PHI (2)	EP (3)	EI (4)	GD (5)
Gains from trade (aggregated)	-0.00142***	-0.00103***	-0.000288**	-0.00113***	-0.0707**
	(0.000504)	(0.000366)	(0.000129)	(0.000307)	(0.0330)
Log number of ethnicities	✓	✓	✓	✓	✓
Ethnic inequality in population	✓	✓	✓	✓	✓
Spatial inequality	✓	✓	✓	✓	✓
Log land Area	✓	✓	✓	✓	✓
Log population,	✓	✓	✓	✓	✓
Region Fixed Effects	✓	✓	✓	✓	✓
N	119	119	119	119	119
R^2	0.672	0.251	0.284	0.806	0.946
Dependent Variable Mean	0.517	0.129	0.0474	0.582	712.3

Note: The unit of observation is a country. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade (aggregated) takes the predicted values of column 1, table F1 (based on equation 6), and takes the country-mean of this value. This is matched to the replication data from Alesina et al. (2016). All outcomes and controls are as described in Alesina et al. (2016).

We follow Alesina et al. (2016) to compare to the following main diversity measures in the literature: ethnolinguistic fractionalization in Alesina et al. (2003a) (henceforth ELF); the peripheral heterogeneity index in Desmet et al. (2009) (henceforth PHI); ethnic polarization from Reynal-Querol and Montalvo (2005) (henceforth EP); genetic diversity in Ashraf and Galor (2013b) (henceforth GD); and finally the aforementioned ethnic inequality measure, from Alesina et al. (2016) (henceforth EI). The results from the exercise can be seen in table F2. In short every one of the main diversity measures behaves in the same manner as our loanwords measure. All five measures are negatively and precisely correlated with gains from trade.

F.3. Migration

Migration is a tricky issue in this context, as the theoretical predictions are ambiguous. On the one hand, reductions in cultural distance should be expected to come along with more migration, as the cultural cost to living in another society is similarly reduced as the transaction costs of inter-cultural trade are reduced. On the other hand, we might expect more migration between geographically homogenous regions because production would be easier in the new location (Michalopoulos, 2012). Since trade partners are unlikely to produce the same things, we might expect land complementarity to be negatively associated with migration.

Thus, the issue is an empirical question. To address it we use the Afrobarometer data, round 6, to assess the share of society i living within the boundaries of society j for all ij pairs.[79] Of course this limits us to African data, but nevertheless provides us with over 11,000 observations of ij-location tuples. We are able to therefore run specifications that include society-relationship fixed effects, or aggregate the data to the societal level for an analogue of our language borrowing analysis that also takes place at that level. We run regressions analogous to our main specification, but investigating mean migration as a dependent variable rather than language exchange. The one deviation that we make is to include a distance from the village-location to the neighbour, which is only relevant since we now have villages within societal boundaries.

Table F3: Migration

	Migrant share of group j within group i					
	location-i-j level		i-j level		society-level	
Utility measure:	percentage change (1)	percentile rank (2)	percentage change (3)	percentile rank (4)	(5)	(6)
Trade influence	-0.00460 (0.00358)	0.0200 (0.0293)	0.00553 (0.00582)	-0.0224 (0.0284)	0.000292 (0.00461)	0.0423 (0.0500)
Gains from trade					-0.00792 (0.00818)	-0.0647 (0.0579)
N	11,162	11,225	670	670	317	317
R^2	0.306	0.179	0.541	0.541	0.199	0.203
Dependent Variable Mean	0.00832	0.00862	0.0207	0.0207	0.0413	0.0413

Note: The unit of observation is either a society pair (i and j) or a society (i) or a location-society pair (location-i-j). *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours. Migrant share is simply share of society i living within the boundaries of society j, with boundaries as defined in Murdock (1959) and current locations as defined in the Afrobarometer.

The results are in table F3, and they are quite inconclusive. We find mostly (but not ubiquitously) positive point estimates for the effect of trade influence on in-migration, using either the raw or percentile rank data. We run the analysis at the i-j-location level (columns 1 and 2), the i-j level (columns 3 and 4) and the i level (columns 5 and 6). We would have expected that lowered cultural barriers to entry would lead to more migration, however this effect cold be attenuated by the (Michalopoulos, 2012) effect, so this is likely what is driving the ambiguity.

F.4. Conflict

One might be concerned that incentive to trade is related to conflict, and through conflict and conquest - rather than strategic investments - language adoption takes place. To investigate this we turned to the PRIO dataset which geocodes conflict events from 1946-2008. We spatially matched conflict events to language groups, and investigated the extent of conflict within a regions boundaries. Taking deaths as a measure of conflict intensity, we are interested in both whether conquest is

[79] Boundaries are determined as in Murdock (1959).

associated with more borrowing as well as whether gains from trade is associated with more conflict. We are motivated by the fact that for conflict to be an alternate explanation for our results it would need to be positively correlated with both gains from trade and language exchange.

Table F4: Conflict

Outcome	Conflict			
Utility measure:			percentage change	percentile rank
	(1)	(2)	(3)	(4)
Language Borrowed	-245.2			
	(174.9)			
Language Loaned		24.38		
		(109.5)		
Gains from trade with neighbours			1,496	11,957
			(1,011)	(10,043)
Influence on trade with neighbours			-205.7	-10,287
			(997.9)	(13,471)
N	2,601	2,601	2,601	2,601
R^2	0.012	0.012	0.013	0.014
Dependent Variable Mean	2429	2429	2429	2429

Note: Conflict data comes from PRIO, and is constructed as the midpoint between the PRIO high and low death estimate, summed over years. *** p<0.01, ** p<0.05, * p<0.1. Gains from trade with neighbours is the percentile rank in the distribution (range [0,1]) of c_i as defined in equation 5, and analogously, influence on trade with neighbours is ι_i. In each case, to aggregate to the societal level we take the maximum value from the society's neighbours.

The results are in table F4. We first look at the conditional correlation between borrowing and lending. There is no attempt at identification here. However, we suspect that either more economic co-operation generates less conflict, or that more gains from trade also means that there is more gains from conquest.

Columns 1 of table F4 strongly suggest the latter. If borrowing were the result of conquest we might expect a positive correlation between conflict and language adoption, instead we see a negative correlation, indicating that more economic integration and contact is associated with less conflict, not more. Column 2 in contrast shows that conflict is unrelated to language lending. In columns 3 and 4 we look at whether our gains from trade or trade influence is associated with conflict. Here we find that neither one is meaningfully associated with conflict, either using the raw values (column 3) or the percentile rank (column 4). We conclude that conflict is unlikely to be driving our results.