

# Language barriers to foreign trade: evidence from translation costs

Alejandro Molnar

World Bank (DECRG)

October 31, 2019  
Macro, Trade and Finance Seminar @ WB

# Language is a factor of production in foreign trade

Knowledge of foreign languages is used intensively in international trade.

Standard, off-the-shelf, gravity estimates usually includes a dummy for *common official language*.

A dummy only measures an average extensive margin for the use of foreign language, and not the intensity or cost of language barriers that need to be overcome.

## This paper:

- Provides a new country-and-language-specific measure of language skill premia.
- Estimates the effect of language cost barriers on trade patterns, and quantifies some of the channels.
- Re-assess the quantitative relevance of language skill abundance or scarcity.  
Macro punchline: Language matters  $\sim 17\times$  more than previously estimated.

# Prior work

## Language barriers to trade

- Melitz & Toubal (JIE, 2014); Fidrmuc & Fidrmuc (EE, 2016): Shared language populations explain trade flows.
- Egger & Lassman (EJ, 2015): Swiss internal languages explain postcode-level exports.
  - Confound barriers and ethnic preferences.
- Papers regressing linguistic distance metrics on trade.
  - e.g. WALs (Lehmann, 2011), ASJP (Isphording & Otten, several), Fearon (2003).
    - Vary at language-pair level. Not structural, no macro impact.
- Ku & Zussman (JEBO, 2010): regresses TOEFL scores on trade → No bilateral variation.

## ... in specialized settings

- Brynjolfsson et al. (MgtSci, 2019): translation fidelity drives cross-border eBay sales.
- Deltas & Evenett (2019): Georgian procurement in English attracts more bidders (RDD).

**Communication and trade:** Fink, Mattoo & Neagu (2005), Freund & Weinhold (2004).

**Returns to language:** Huge literature. e.g. Altonji '95, Bleakley & Chin '04, Levinsohn '07, Albouy '08, Chiswick & Miller '10, Shastry '12, Stöhr '15, Chakraborty and Bakshi '16.

**Language 'structure' on behavior:** Jakiela & Ozier (2019), Chen (2013).

# Translation cost data from [translatorscafe.com](https://translatorscafe.com)

- Online markets for translation services provide a measure of language skill premia.
  - ▶ Translators submit a *private* rate at which the platform screens jobs.
  - ▶ Platform reports this average reservation wage by language pair and translator's country of residence.
- Within a language pair, rates vary between countries because:
  - ▶ Translation may require local knowledge (e.g. culture, laws).
  - ▶ Translators seek work online to fill to capacity. Local, offline wage determines the opportunity cost of time.
- Example data: translation rates per word from English into  $X$  for translators located in:

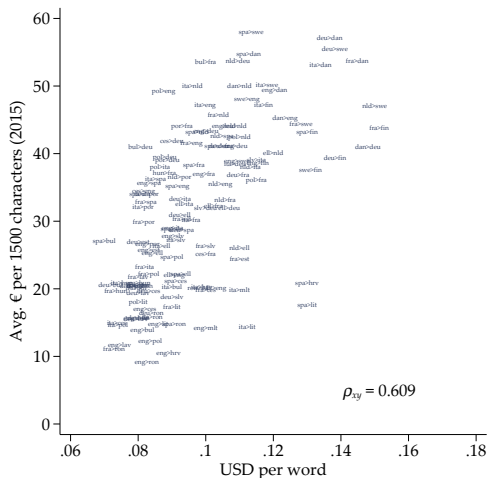
For  $X$  = French

- ▶ France 0.10 USD
- ▶ Côte d'Ivoire: 0.12 USD
- ▶ Morocco: 0.10 USD
- ▶ Senegal: 0.08 USD

For  $X$  = Spanish

- ▶ Spain 0.08 USD
- ▶ Guatemala: 0.10 USD
- ▶ Paraguay: 0.09 USD
- ▶ Uruguay: 0.08 USD

# Translation cost data: validation



← Procurement. Transaction data from the European Union Court of Justice (2015)

▶ Compare against online rates averaged over EU translators.

● Procurement. Contract price sheets from the US General Services Administration

▶ US CSA

● Surveys of translator rates conducted by four national associations

▶ USA, GBR, FRA, DEU

▶ Assoc.

# Translation cost data

## Pros

- Plausibly exogenous in some applications, e.g. destinations at the firm-level.
- Varies by country-pair, not just countries or language pairs.
- Asymmetric (may differ at origin vs. destination).
- Skill premia are a relevant price for structural trade models.

## Cons

- Endogenous at country-pair level, e.g. gravity.
- Missing in potential language pairs (e.g. Hindi-German in Germany)
- Variation in unit of account.
- Only a cross-section.

# Translation cost data: construction

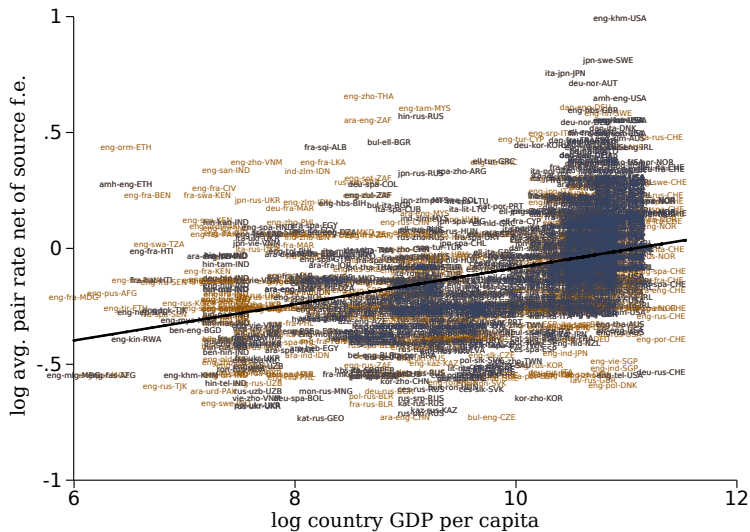
Goal: measure resource cost of a language-based task required for trade by country pairs.

For translation rate  $r_{stc}$  between source language  $s$  to target language  $t$  for translators located in country  $c$ , let:

$$r_{stc} = \delta_s \times \delta_{(s,t)c} \times w_c \times \eta_{stc}$$

- Source language f.e. control for unit of account.
- Net of local GDP per capita to obtain skill premia relative to wages.
- Rates will be normalized relative to the English-Spanish skill premium in the US.

# Translation cost data

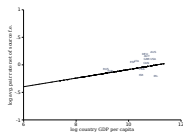


Language pairs including the country's top spoken language (blue) or not (orange).

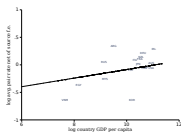


# Translation cost data

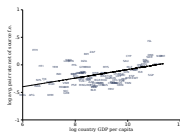
Country level translation costs between majority language and ...



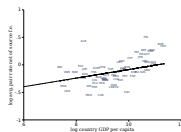
Arabic



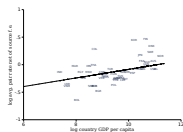
Chinese



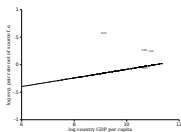
English



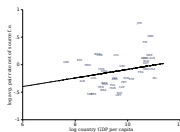
French



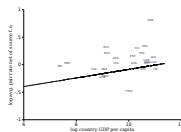
German



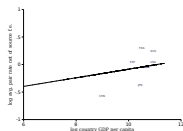
Hindi



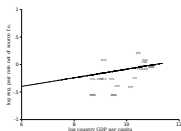
Italian



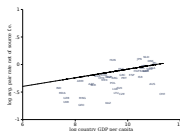
Japanese



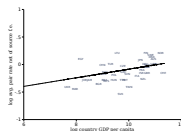
Korean



Portuguese



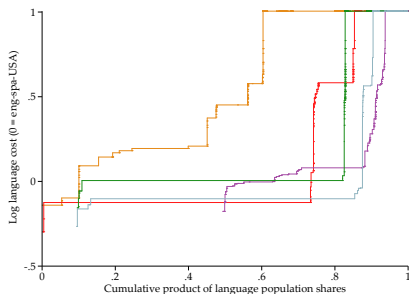
Russian



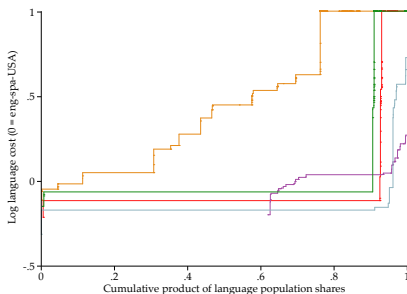
Spanish

# Construction of bilateral language cost

- Pop. of languages in country from *Ethnologue* (various editions).
- Language population cross-product cells ordered by avg. bilateral language cost within cell, for each of 5 partners.
- Missing are censored at  $\sim 1.05$ , highest observed log-rate.
- Prices are normalized at  $\log lc_{\text{eng-spa-USA}} = 0$ .
- Country-pair measure is the ordered, cell-weighted average.



United States

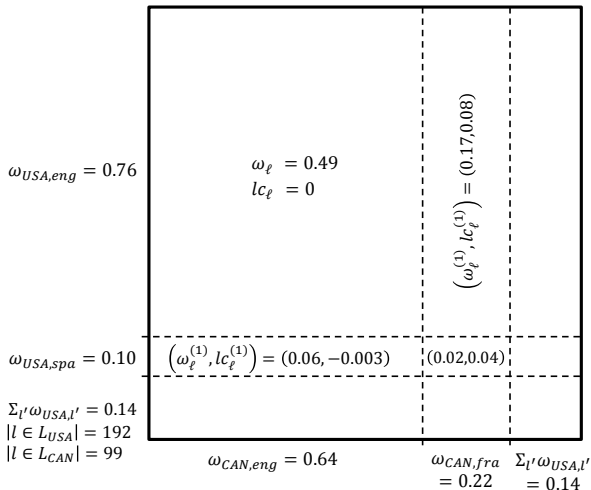


United Kingdom

Partner colors are: India, China, Mexico, Argentina, Canada.

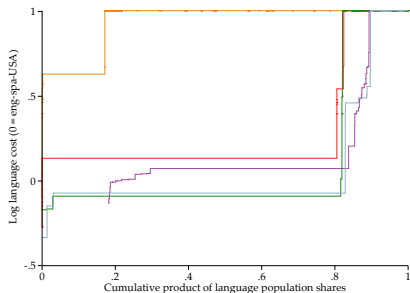
# Construction of bilateral language cost

Example: language population cells for USA-Canada

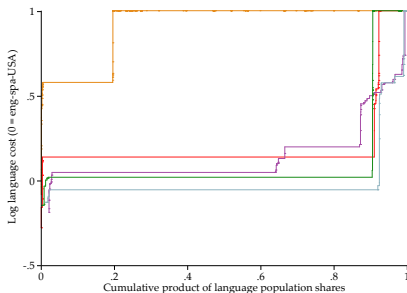


# Construction of bilateral language cost

- Language population cross-product cells ordered by avg. bilateral language cost within cell, for each of 5 partners.
- Missing are censored at  $\sim 1.05$ , highest observed log-rate.
- Prices are normalized at  $\log lc_{\text{eng-spa-USA}} = 0$ .
- Country-pair measure is the ordered, cell-weighted average.



France

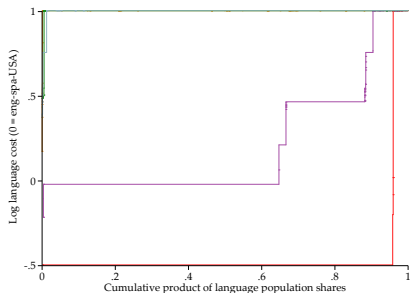


Germany

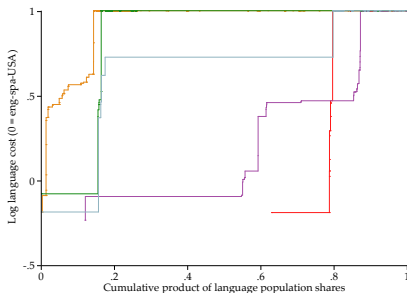
Partner colors are: India, China, Mexico, Argentina, Canada.

# Construction of bilateral language cost

- Language population cross-product cells ordered by avg. bilateral language cost within cell, for each of 5 partners.
- Missing are censored at  $\sim 1.05$ , highest observed log-rate.
- Prices are normalized at  $\log lc_{\text{eng-spa-USA}} = 0$ .
- Country-pair measure is the ordered, cell-weighted average.



South Korea



Singapore

Partner colors are: India, China, Mexico, Argentina, Canada.

# A baseline gravity regression

## Data

- Standard geographical gravity covariates from CEPII.
- Trade flows from BACI. All trade 2003-2016.
- 205 countries, no internal trade. At most 41820 ( $= 205 \times 204$ )  $ij$  pairs.

## Estimating equation

$$\log X_{ij} = \beta_c \text{Language cost}_{ij} + \beta_s \text{Same language share}_{ij} \\ + \text{Gravity covariates}_{ij} + \varepsilon_{ij}$$

## Where

- $LC_{ij} = \sum_{ll'} \omega_{ll'} (0.5 lc_{ill'} + 0.5 lc_{jll'})$  for  $l \in \mathcal{L}_i$  and  $l' \in \mathcal{L}_j$ , and  
when  $\sum_{ll'} \omega_{ll'} > 1$ , weights  $\omega_{ll'}$  are included in the order induced by  $lc_{ll'}$ .
- $LS_{ij} = \sum_{ll'} \omega_{ll'} \mathbb{1}(l \neq l')$
- All specifications include origin & destination f.e.

# Estimation strategy for language cost in gravity

- Endogeneity of skill-premia:

- ▶ Trade raises demand for and price of translations, attenuating the estimated effect of premium on trade.
- ▶ Low translation cost can be correlated with shared ethnicity, causing trade through an ethnic, non-language channel

- **IV strategy #1:** language distance measures.

Alternatives: **ASJP** (phonetic), **WALS** (phonetic+grammar+lexicon), language tree “**cleavages**” (e.g. Fearon, 2003)

- **IV strategy #2:** overlaps in world language population

- ▶ The populations that speak some languages overlap more than others (e.g. English and Vietnamese, more than English and Thai).
- ▶ This will have a general equilibrium effect on language premia.
- ▶ **Overlap IV:** Excluded ethnolinguistic overlap (i.e. probability of national co-habitation for language speakers from the country-pair, excluding the pair).

- Limitations of all of these IVs:

1. Coarsen true variation to the language-pair.
2. Symmetric.

# Gravity linear regression. Positive trade 2003-2016

Table I: Gravity estimation. Linear conditional mean on sample with positive flows over 2003-2016.

	OLS					2SLS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Language cost			-0.185 <sup>a</sup> (0.064)	-1.888 <sup>a</sup> (0.200)	-1.413 <sup>a</sup> (0.151)	-1.500 <sup>a</sup> (0.189)	-1.377 <sup>a</sup> (0.149)	-1.321 <sup>a</sup> (0.281)	-1.635 <sup>a</sup> (0.185)	-1.653 <sup>b</sup> (0.688)	-1.876 <sup>a</sup> (0.196)
Same lang. share		0.446 <sup>a</sup> (0.095)	0.328 <sup>a</sup> (0.103)	-0.756 <sup>a</sup> (0.156)	-0.454 <sup>a</sup> (0.134)	-0.509 <sup>a</sup> (0.150)	-0.431 <sup>a</sup> (0.133)	-0.395 <sup>b</sup> (0.197)	-0.595 <sup>a</sup> (0.152)	-0.607 (0.448)	-0.749 <sup>a</sup> (0.153)
Common off. lang.	0.798 <sup>a</sup> (0.039)	0.708 <sup>a</sup> (0.046)	0.717 <sup>a</sup> (0.046)	0.802 <sup>a</sup> (0.047)	0.778 <sup>a</sup> (0.047)	0.783 <sup>a</sup> (0.047)	0.777 <sup>a</sup> (0.047)	0.774 <sup>a</sup> (0.047)	0.789 <sup>a</sup> (0.047)	0.790 <sup>a</sup> (0.058)	0.801 <sup>a</sup> (0.047)
Log distance	-1.503 <sup>a</sup> (0.018)	-1.488 <sup>a</sup> (0.018)	-1.484 <sup>a</sup> (0.018)	-1.443 <sup>a</sup> (0.019)	-1.455 <sup>a</sup> (0.018)	-1.452 <sup>a</sup> (0.019)	-1.455 <sup>a</sup> (0.018)	-1.457 <sup>a</sup> (0.019)	-1.449 <sup>a</sup> (0.018)	-1.449 <sup>a</sup> (0.025)	-1.444 <sup>a</sup> (0.019)
Contiguity	0.783 <sup>a</sup> (0.104)	0.768 <sup>a</sup> (0.104)	0.746 <sup>a</sup> (0.105)	0.548 <sup>a</sup> (0.110)	0.603 <sup>a</sup> (0.109)	0.593 <sup>a</sup> (0.109)	0.607 <sup>a</sup> (0.108)	0.614 <sup>a</sup> (0.108)	0.577 <sup>a</sup> (0.109)	0.575 <sup>a</sup> (0.136)	0.549 <sup>a</sup> (0.110)
Colonial tie (ever)	0.062 (0.107)	0.018 (0.106)	0.021 (0.107)	0.055 (0.115)	0.045 (0.112)	0.047 (0.112)	0.045 (0.111)	0.044 (0.111)	0.050 (0.113)	0.050 (0.114)	0.055 (0.115)
Colonial tie (after 1945)	1.217 <sup>a</sup> (0.150)	1.262 <sup>a</sup> (0.150)	1.274 <sup>a</sup> (0.151)	1.380 <sup>a</sup> (0.163)	1.350 <sup>a</sup> (0.159)	1.356 <sup>a</sup> (0.160)	1.348 <sup>a</sup> (0.158)	1.344 <sup>a</sup> (0.159)	1.364 <sup>a</sup> (0.161)	1.365 <sup>a</sup> (0.165)	1.379 <sup>a</sup> (0.163)
Common colonizer (after 1945)	0.566 <sup>a</sup> (0.049)	0.591 <sup>a</sup> (0.049)	0.585 <sup>a</sup> (0.049)	0.525 <sup>a</sup> (0.050)	0.542 <sup>a</sup> (0.050)	0.539 <sup>a</sup> (0.050)	0.543 <sup>a</sup> (0.050)	0.545 <sup>a</sup> (0.050)	0.534 <sup>a</sup> (0.050)	0.533 <sup>a</sup> (0.055)	0.525 <sup>a</sup> (0.050)
Instruments				ASJP	WALS <sub>1</sub>	WALS <sub>2</sub>	WALS <sub>1,2</sub>	AP <sub>1,2</sub>	Fearon <sub>1-14</sub>	Overlaps	ASJP & overlaps
Observations	34014	34014	34014	34014	34014	34014	34014	34014	34014	34014	34014
R <sup>2</sup>	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
J-test p-value				0.58	0.013	0.0059	0.016	0.016	6.1e-10		0.80
1st Stg. F (KP)				4208.4	5754.1	4325.4	3893.9	841.2	320.5	11.9	2841.8



# Gravity exponential regression. All trade 2003-2016

Table IV: Gravity GMM estimation with exponential conditional mean. 2003-2016.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Language cost	-1.250 <sup>a</sup> (0.214)	-0.455 <sup>b</sup> (0.188)	-1.332 <sup>a</sup> (0.438)	-0.426 <sup>b</sup> (0.178)	-2.229 (2.663)	-1.288 <sup>a</sup> (0.214)	-0.422 <sup>b</sup> (0.179)	-0.762 <sup>a</sup> (0.131)
Same lang. share	0.375 (0.238)	0.894 <sup>a</sup> (0.203)	0.339 (0.244)	0.392 <sup>c</sup> (0.204)	0.251 (0.286)	0.379 (0.239)	0.814 <sup>a</sup> (0.201)	0.485 <sup>b</sup> (0.193)
Common off. lang.	-0.164 (0.109)	-0.354 <sup>a</sup> (0.110)	-0.161 (0.111)	-0.134 (0.110)	-0.168 (0.153)	-0.118 (0.109)	-0.234 <sup>b</sup> (0.104)	-0.141 (0.106)
Log distance	-0.557 <sup>a</sup> (0.044)	-0.734 <sup>a</sup> (0.038)	-0.545 <sup>a</sup> (0.061)	-0.676 <sup>a</sup> (0.034)	-0.415 (0.291)	-0.543 <sup>a</sup> (0.043)	-0.717 <sup>a</sup> (0.034)	-0.643 <sup>a</sup> (0.031)
Contiguity	0.600 <sup>a</sup> (0.104)	0.417 <sup>a</sup> (0.090)	0.656 <sup>a</sup> (0.103)	0.576 <sup>a</sup> (0.073)	0.694 <sup>b</sup> (0.280)	0.587 <sup>a</sup> (0.106)	0.383 <sup>a</sup> (0.087)	0.596 <sup>a</sup> (0.077)
Colonial tie (ever)	0.051 (0.101)	0.050 (0.107)	0.081 (0.101)	0.163 <sup>c</sup> (0.094)	0.013 (0.124)	0.039 (0.100)	0.013 (0.108)	0.121 (0.093)
Colonial tie (after 1945)	0.051 (0.232)	0.986 <sup>a</sup> (0.263)	-0.048 (0.242)	0.041 (0.208)	-0.074 (0.461)	0.059 (0.229)	0.806 <sup>a</sup> (0.242)	-0.010 (0.219)
Common colonizer (after 1945)	0.181 (0.141)	0.297 <sup>c</sup> (0.162)	0.129 (0.149)	0.103 (0.147)	0.196 (0.178)	0.170 (0.139)	0.271 <sup>c</sup> (0.161)	0.027 (0.145)
Instruments	ASJP	WALS <sub>1,2</sub>	AP <sub>1,2</sub>	Fearon <sub>1-8</sub>	Overlaps	ASJP & overlaps	ASJP, Over. & WALS <sub>1,2</sub>	ASJP, Over. & Fearon <sub>1-8</sub>
Observations	41820	41820	41820	41820	41820	41820	41820	41820

# Robustness

- Alternative language cost construction:  
e.g., screen for # of translators in sample; functional form of gdp regressor
- Editions of *Ethnologue* language population data.
  - ▶ Baseline is 16<sup>th</sup> edition, which I will distribute.
  - ▶ Robust to 20<sup>th</sup> edition.
    - L1 speakers only.
    - Includes L2 data, but coverage is inconsistent.

Note: 20<sup>th</sup> edition has prohibitive license and missing citation-years.
- Alternative trade flow years, e.g. each of 2003-2016.
- Substantive econometric concern #1: robust to alternative instruments.

# Missingness

Substantive econometric concern #2: missing translation cost data.

Observed status:	Neither	Exporter only	Importer only	Both	Same language	Total
Raw count	10,748,297	55,105	55,107	13,067	37,766	10,909,342
Pop. weighted	33,762.2	2,887.4	2,887.4	1,164.6	1,118.4	41,820
Trade weighted	31.0%	6.4%	5.9%	47.9%	8.8%	100%

Use two approaches:

- 1 i) Assume censoring. ii) Control function approach of Chernozhukov, Rigobon and Stoker (2010). [▶ Chernozhukov et al.](#)
- 2 i) Define the language cross-product cell as the observation, and impute missing data. ii) Adapt a GMM imputation method due to Abrevaya and Donald (2014) to bilateral data and two-sided missingness.

[▶ Abrevaya-Donald GMM](#)

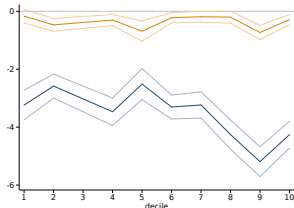
# Counterfactuals: partial effects and Full GE

(Anderson, Larch and Yotov, 2018)

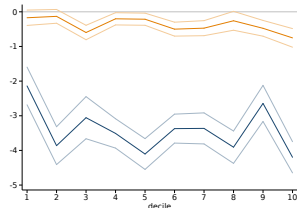
Counterfactual	Partial effect		Full GE counterfactual			
	Trade flow ratio (avg.)	Relative to common lang.	Change in total world trade		Relative to common lang.	
Eliminate language barriers (comm. lang. measure)	1.059		1.029			
Eliminate language barriers (lang. cost measure)	4.237	54.5	1.507	[1.495,1.578]	17.6	[17.5,18.0]
Eliminate distance	31.900	520.2	3.495	[3.359,4.258]	86.4	[83.7,101.2]

- For full GE, construct domestic shares from CEPII TradeProd database.
- Re-estimate trade cost geographic elasticities with domestic shares and an external border effect.
- Result: in GE, eliminating language barriers increases trade by 50.7%.
  - ▶ With common language only, effect of language is only 2.9%.
  - ▶ As a benchmark, eliminating distance increases GE trade by 249.5%.

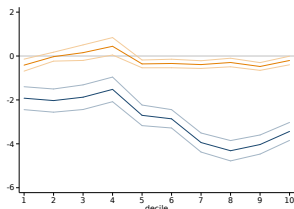
# Language cost elasticity by product attribute



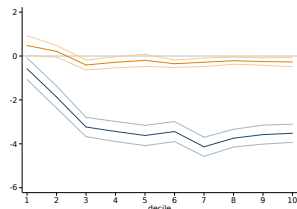
R&D intensity (Kugler and Verhoogen, 2012)



Adv. intensity (Kugler and Verhoogen, 2012)



Contract intensity (Nunn, 2007)



"Product complexity" (Hidalgo et al., 2007)

Plots coefficient estimates for subsample of HS6 codes by attribute decile. **PPML** or **COP-2SLS** on HS6 subsamples by attribute decile.

# Language cost elasticity by HS2 product category

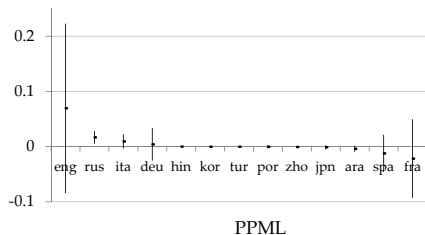
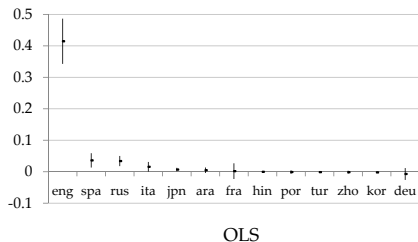
$\hat{\eta}_{lc}^h$	$\hat{\sigma}_{lc}^h$	HS2	Description
-5.414	0.398	30	Pharmaceutical products
-4.704	0.423	29	Organic chemicals
-4.521	0.874	93	Arms and ammunition; parts and accessories thereof
-4.410	0.380	34	Soap, organic surface-active agents
-4.335	0.349	33	Essential oils and resinoids; perfumery, cosmetic or toilet preparations
-4.174	0.313	39	Plastics and articles thereof
-4.134	0.354	22	Beverages, spirits and vinegar
-3.999	0.486	04	Dairy produce; birds eggs; natural honey
-3.857	0.497	72	Iron and steel
-3.746	0.360	48	Paper and paperboard; articles of paper pulp, of paper or of paperboard
0.349	0.509	06	Live trees and other plants
-0.468	0.303	63	Other made up textile articles; sets; worn clothing and worn textile articles
-0.697	0.522	46	Manufactures of straw, of esparto or of other plaiting materials
-0.700	0.482	11	Products of the milling industry; malt; starches; inulin
-0.871	0.524	18	Cocoa and cocoa preparations
-0.878	0.480	16	Preparations of meat, of fish or of crustaceans
-0.899	0.393	07	Edible vegetables and certain roots and tubers
-1.243	0.435	52	Cotton
-1.261	0.655	50	Silk
-1.273	0.428	13	Lac; gums, resins and other vegetable saps and extracts

## Firm-level evidence (preliminary)

- Replicate Fernandes et al. (2016) study of WB Exporter Dynamics Database.
- Language cost operates mostly on the extensive margin.
- OLS evidence that language cost matters more on the exporter side.

## *Lingua franca* impact, by language (preliminary)

Specification: jointly include a product measure of language cost through alternative third languages.



Caveat: this exercise is not entirely coherent yet, but ...  
clearly the only major outlier is English.

# Conclusions

- First measure of language-specific skill premium by country.
- Empirical strategy to identify effect of language barriers, separate from ethnic trade.
  - ▶ Properly measured, language barriers play a substantially larger role in trade patterns than previously identified.
- Impact of language barriers increases in measures of differentiation, product R&D, Ad spend, contract-intensity, “complexity”.
- Evidence that language barriers operate strongly through extensive margin.
- Evidence that English plays a unique role as a lingua franca



Extra stuff

Figure: Translation rates. US GSA procurement

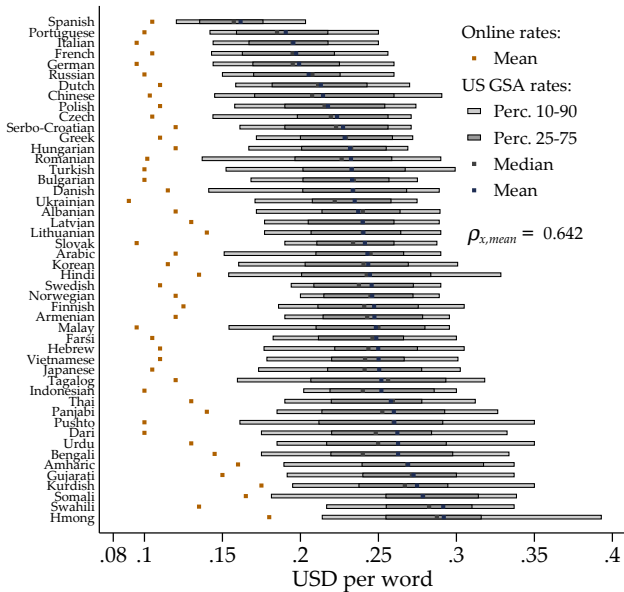
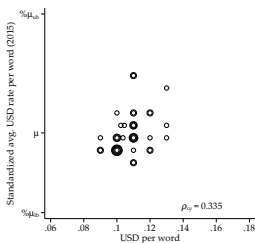
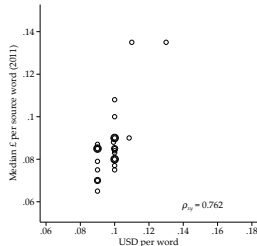


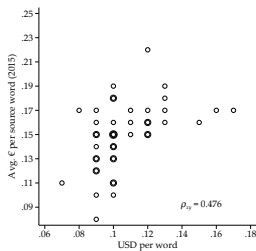
Figure: Translation rates. Survey data from national translator associations



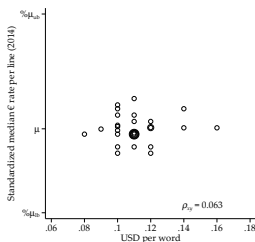
(a) United States (ATA, 2014)



(b) United Kingdom (ITI, 2011)



(c) France (STF, 2015)



(d) Germany (KBDU, 2014)

► back

# Missing translation rates.

## Chernozhukov, Rigobon and Stoker (2010) control function approach

Table B.I: Control function gravity estimates (Chernozhukov *et al.*, 2010).

	(1)	(2)	(3)	(4)	(5)	(6)
Language cost	-1.484 <sup>a</sup> (0.176)	-1.257 <sup>a</sup> (0.145)	-0.881 <sup>a</sup> (0.206)	-1.022 <sup>a</sup> (0.146)	-0.550 <sup>a</sup> (0.211)	-1.455 <sup>a</sup> (0.170)
$\hat{V}^*$	1.520 <sup>a</sup> (0.187)	1.386 <sup>a</sup> (0.162)	0.764 <sup>a</sup> (0.212)	1.032 <sup>a</sup> (0.155)	0.372 <sup>c</sup> (0.203)	1.495 <sup>a</sup> (0.181)
Same lang. share	-0.484 <sup>a</sup> (0.143)	-0.461 <sup>a</sup> (0.138)	-0.286 (0.184)	-0.357 <sup>b</sup> (0.143)	-0.042 (0.191)	-0.502 <sup>a</sup> (0.143)
Common off. lang.	0.767 <sup>a</sup> (0.047)	0.787 <sup>a</sup> (0.047)	0.823 <sup>a</sup> (0.048)	0.848 <sup>a</sup> (0.048)	0.810 <sup>a</sup> (0.049)	0.773 <sup>a</sup> (0.047)
Log distance	-1.458 <sup>a</sup> (0.018)	-1.459 <sup>a</sup> (0.018)	-1.478 <sup>a</sup> (0.018)	-1.468 <sup>a</sup> (0.018)	-1.479 <sup>a</sup> (0.018)	-1.456 <sup>a</sup> (0.018)
Contiguity	0.403 <sup>a</sup> (0.090)	0.490 <sup>a</sup> (0.089)	0.436 <sup>a</sup> (0.093)	0.450 <sup>a</sup> (0.090)	0.472 <sup>a</sup> (0.094)	0.421 <sup>a</sup> (0.090)
Colonial tie (ever)	0.106 (0.154)	0.141 (0.154)	0.131 (0.154)	0.098 (0.154)	0.111 (0.154)	0.112 (0.154)
Colonial tie (after 1945)	1.269 <sup>a</sup> (0.196)	1.233 <sup>a</sup> (0.195)	1.210 <sup>a</sup> (0.196)	1.282 <sup>a</sup> (0.196)	1.200 <sup>a</sup> (0.196)	1.281 <sup>a</sup> (0.196)
Common colonizer (after 1945)	0.555 <sup>a</sup> (0.050)	0.571 <sup>a</sup> (0.050)	0.559 <sup>a</sup> (0.050)	0.544 <sup>a</sup> (0.050)	0.571 <sup>a</sup> (0.050)	0.557 <sup>a</sup> (0.050)
Instruments	ASJP	WALS <sub>1,2</sub>	AP <sub>1,2</sub>	Fearon <sub>1-14</sub>	Overlaps	ASJP & overlaps
Observations	24838	24838	24838	24838	24838	24838
$R^2$	0.82	0.82	0.82	0.82	0.82	0.82

# Missing translation rates.

## Abrevaya and Donald (2016) GMM imputation

Table C.III: GMM imputation estimates, following [Abrevaya and Donald \(2016\)](#)

	Gravity equation		Language cost projection	
Language cost	-15.95	[-20.86,-15.49]		
Same language	1.350	[-3.006, 1.557]		
<b>Included instruments</b>	$\delta$		$\gamma_2$	
Common official language	-1.989	[-3.295, 1.985]	-0.172	[-0.265, 0.068]
log Distance	-2.134	[-2.436,-1.987]	-0.080	[-0.090,-0.067]
Contiguity	-0.351	[-2.838,-0.215]	-0.130	[-0.232,-0.103]
Colonial tie (ever)	3.481	[ 1.347, 3.552]	0.029	[-0.081,0.113]
Colonial tie (after 1945)	10.613	[ 9.660,10.651]	0.757	[ 0.567,0.825]
Common colonizer (after 1954)	0.176	[-0.929, 0.950]	0.046	[-0.024,0.077]
log GDP and remoteness (o&d)	Y		Y	
<b>Excluded instruments</b>			$\gamma_1$	
Linguistic distance (ASJP)			0.003	[0.002,0.004]
Linguistic distance missing			0.417	[0.296,0.496]
Ethnolinguistic overlap			-136.2	[-136.3,-135.2]