

Data

Language

CHELEM

Comptes Harmonisés sur les
Echanges et L'Economie
Mondiale (CHELEM)

Geography

GeoDist
Gravity
Language
Market Potentials

Macroeconomics

EconMap
EQCHANGE
Institutional Profiles
Séries longues
macroéconomiques

Production &
Specialization Indicators

Export Sophistication
ProdComp
TradeProd

Profiles

Profils Pays / Country Profiles

Trade & International
Investments

BACI
FDIMap
Network Trade
Product Level Trade Elasticities
RCA
Trade Unit Values (TUV)
TradePrices
TRADHIST
World Trade Flows
Characterization

Trade Protection

AVEs-Services
MacMap-HS6
NTM-MAP

Models

Long Term Growth

MaGE

Trade Policy

MIRAGE

Language provides new series for Common Official Language (COL), Common Spoken Language (CSL), Common Native Language (CNL) and in the case of Linguistic Proximity, LP1 and LP2, the unadjusted values that serve for constructing two different measures, which we label PROX1 and PROX2. This series are available for 195 countries.

A do-file will permit you to construct our series for LP1 and LP2. But it will also permit you to construct both variables based on your own dataset. This is important in the case of LP1 and LP2 since both variables are data dependent. The same do-file will permit you to construct a unique Common Language index (CL) as well, a variable which depends, in turn, on LP and therefore is also data dependent.

Methodology: A Short Description

In this data, as everywhere, this measure of Common Official Language (COL) is a binary one, either 0 or 1. With regard to COL, the usual source is the *CIA World Factbook*. Though we used it as well, we adopted a slightly broader definition of COL.

With regard to Common Native Language (CNL) and Common Spoken Language (CSL), we required all languages to be spoken by at least 4% of the population in 2 countries. The data on native and spoken language are collected from various sources described in detail in the paper.

We constructed two separate measures of Linguistic Proximity, LP1 and LP2. LP1 is inspired by the idea in Fearon (2003) and Laitin (2000) of calculating linguistic proximities on the basis of the *Ethnologue* classification of language trees between trees, branches and sub-branches. As regards LP2, the source is an analysis of lexical similarity between 40 words that were compiled by the Automated Similarity Judgment Program (ASJP) project.

Presentation

Reference document to mention

Jacques Melitz and Farid Toubal, 2012. **Native language, spoken language, translation and trade**. CEPII, Working Papers 2012-17.

Reference document to mention

Native Language, Spoken Language, Translation and Trade Journal of International Economics, Vol. 92, N°2, 2014 , p. 351-363.

Jacques Melitz , Farid Toubal

Download Page

johnson@demogr.mpg.de

ok

Register

Person in charge

Jacques Melitz & Farid Toubal

Contact

ftoubal@ens-cachan.fr



LICENCE OUVERTE
OPEN LICENCE

Follow @CEPII_Paris

Tweet

Share

