Linguistic Fractionalization and Health Information in Sub-Saharan Africa

Joseph Flavian Gomes
IRES/LIDAM, UCLouvain, (email:joseph.gomes@uclouvain.be)

October 2018

Abstract

This paper explores the relationship between linguistic diversity and the stock of health information in society. Information is measured using individual-level knowledge about the oral rehydration product for treating children with diarrhoea. Exploiting an individual woman-level dataset from the Demographic and Health Surveys for 14 sub-Saharan African countries combined with a novel high-resolution dataset on the spatial distribution of linguistic groups at a 1 km × 1 km level, this study shows that linguistic diversity has an inverted U-shaped relationship with the stock of information in society.

Introduction

The cross-country literature on ethnolinguistic diversity has typically found diversity to have a negative effect on different socio-economic outcomes such as economic development (Easterly and Levine, 1997), provision of public goods (Alesina et al., 1999), and redistribution (Desmet et al., 2009). In sharp contrast, more localized measures of diversity have been shown to have a positive effect on inter-ethnic trust (Robinson, 2013), group creativity (McLeod et al., 1996), city growth (Glaeser et al., 1995; Ottaviano and Peri, 2006), and even the provision of public goods (Desmet et al., 2018).

In recent research, Ashraf and Galor (2013) find a non-linear hump-shaped relationship between diversity and economic development outcomes, highlighting the trade-off between the beneficial and detrimental effects of diversity. Their hypothesis is that higher diversity reflects the existence of more complementary traits in society, expanding its production possibility frontier. However, as diversity keeps going up, it leads to an increase in the possibilities of disarray and mistrust which reduces cooperation and disrupts socioeconomic order. This in turn has a negative effect on the economy (Ashraf and Galor, 2013).

We take this literature forward by showing how diversity might indeed have a non-linear hump-shaped relationship with the stock of information or knowledge in society. We exploit a novel dataset recently constructed by Gomes (2017). This dataset combines individual-level data on more than 205,000 women, from the Demographic and Health Surveys (DHS) for 14 sub-Saharan African countries, together with a new high-resolution dataset on the spatial distribution of linguistic groups at the 1 km × 1 km level.

Based on this novel dataset, Gomes (2017) makes available several different localized measures of linguistic diversity at both the individual and region levels. Gomes

(2017) begins by making a fine distinction between two alternative concepts of diversity. The first is the more commonly used aggregated measure of *linguistic fractionalization* measured at the region level, which is defined as the probability that two randomly selected individuals from a given region speak two different languages. However, his primary focus is on the concept of individual-level *linguistic distance*, which measures how linguistically different an individual is from others living in the same region. While fractionalization is region specific, linguistic distance is ethnicity-region specific.

Using the newly constructed database, Gomes (2017) puts forth two primary findings. First, children of mothers who are linguistically more distant from their neighbours have a higher probability of dying before reaching the age of five. And second, linguistic fractionalization has a more benign (often statistically insignificant) effect on the child health outcomes. His results are robust to the inclusion of several individual specific controls, apart from ethnicity, region, and country-time fixed effects. He argues that linguistic distances act as barriers to health-related knowledge worsening health outcomes.

We take the findings of Gomes (2017) another step forward by focussing on the relationship between diversity and health-related knowledge in society. Following the insights of Ashraf and Galor (2013), we allow for a non-linearity in the relationship between diversity and knowledge in society. In particular, we exploit a question from the DHS which asks all the interviewed women in the sample whether they have heard about the oral rehydration solution (ORS) product for treating children with diarrhoea. Using the individual mother-level data, we find that linguistic fractionalization has a non-linear hump-shaped relationship with the probability of knowing about ORS. At lower levels of diversity, having some more diversity increases access to information. However, at higher levels having yet some more diversity actually reduces information. For instance, if we calculate diversity by considering a circle of 50 km in radius around the mother as the relevant region, diversity continues to increase knowledge about ORS until a diversity level of 0.44, but further increases in diversity reduces such knowledge. These results are robust to the inclusion of the individual-level linguistic distance variable, apart from ethnicity, country and birth cohort fixed effects, among other controls. The linguistic distance variable, on the other hand, reduces the probability of knowing about the ORS product (Gomes, 2017).

Empirical Analysis

The empirical analysis exploits a novel dataset constructed by Gomes (2017). This dataset combines individual-level geo-coded data from the DHS for 14 sub-Saharan African countries with a newly-constructed database on the spatial distribution of linguistic groups at the 1 km × 1 km level. These latter data are based on an iterative proportional fitting algorithm recently developed by Desmet et al. (2018) applied to maps of linguistic groups and population distribution from the Ethnologue and Landscan databases.[1] While Gomes (2017) focuses on the effects of linguistic *distance* on child health, we focus on the effects of linguistic *fractionalization* on health-related information.

We construct our main dependent variable based on a question from the DHS about the respondent's knowledge of the oral rehydration product for treating children with diarrhoea. This lets us create a 0-1 binary variable called ORS which takes the value '1' if the individual has either heard of or used the product and '0' otherwise. Our main independent variable of interest is the commonly used ELF index defined as:

$$ELF(j) = 1 - \sum_{i=1}^{n} S_i\,(j)^2 \tag{1}$$

where $ELF(j)$ is the linguistic fractionalization index for region *'j'*, and $S_i\,(j)$ is the proportion of the population speaking language *'i'* in region *'j'*. *'n'* is the number of languages spoken in the region *'j'*. This fractionalization index $ELF(j)$ is defined as the probability that any two randomly selected individuals from a given region *'j'* speak two different languages. It ranges between zero and one, where one represents maximum fractionalization in society. It is maximized when every individual in the region under consideration speaks a different language.[2]

Our primary econometric specification is given by equation (2):

$$ORS_{iet} = a_{Ethnicity} + a_{Religion} + a_t + a_{country} + \beta_1 ELF + \beta_2 ELF^2 + \beta_3 LD + \beta_4 X_i + \varepsilon_{it} \tag{2}$$

where ORS_{iet} is a 0-1 binary variable showing whether the interviewed woman *'i'* belonging to ethnicity *'e'* born in year *'t'* has either heard of or used the ORS product. *ELF* measures the linguistic fractionalization in the region, and ELF^2 is the square of the *ELF* index. Hence, β_1 and β_2 are our main coefficients of interest. *LD* gives the individual-level linguistic distance of the woman *'i'* belonging to ethnicity *'e'* from people living around her.[3] Following Gomes (2017) we calculate both the *ELF* and *LD* variables in regions constructed by drawing circles of different radii, namely 25, 50, 75, 100, 125, 150, 175, 200, 250 km around the woman. $a_{Ethnicity}$, $a_{Religion}$, a_t, and $a_{country}$ respectively control for the ethnicity, religion, birth cohort and country fixed effects. The other controls X_i include an urban-location dummy, dummies for the wealth index, dummies for the educational attainment of the woman, log of population in the circle (which effectively controls for population density) and the log of distance to the capital. Finally, ε_{it} represents the error term. Since ORS_{iet} is a binary variable, we estimate a linear probability model.

[1] Figure 1 gives the geocoded locations of the individual mothers in the sample.
[2] See Alesina et al. (2003) for a more detailed discussion on the fractionalization index.
[3] The linguistic distance measure is based on language trees from the Ethnologue database. Please see Gomes (2017) for more details.

In Table 1, we provide the summary statistics of the primary variables from our sample. In Table 2, we present the results from our analysis. The dependent variable in each of the three panels of Table 2 is the ORS variable defined above. All specifications include the different controls listed in equation (1). The different columns represent the circles of different radii drawn around the woman ranging from 25 km to 125 km to calculate the ELF, linguistic distance and population variables.[4]

First, in Panel 1 of Table 2, we show how ELF by itself does not have a statistically significant effect on ORS knowledge. However, in Panel 2 when we introduce the ELF^2 term we see that the ELF variable, in fact, has a non-linear effect on information about ORS. While the coefficient on the linear term of ELF has a positive sign, the coefficient on the square term is negative indicating an inverted U-shaped relationship underscored by Ashraf and Galor (2013). We plot this hump-shaped relationship for circles of different radii in Figure 2. We see that an increase in diversity continues to increase knowledge about ORS until a diversity level of 0.44 (considering a circle of 50 km in radius) and then further diversity reduces knowledge.

Finally, in Panel 3, we show that our main results are robust to controlling for the individual linguistic distance variable from Gomes (2017). As highlighted by Gomes (2017), the linguistic distance variable has a negative effect on ORS knowledge. However, our primary variable of interest ELF continues to have a statistically significant hump-shaped effect on ORS.

Conclusion

Using high-quality individual-level data from 14 sub-Saharan African countries we show how linguistic fractionalization has a hump-shaped effect on information about ORS. At low levels of diversity, having some more diversity increases knowledge about ORS, but after a certain level further increases in diversity reduces such knowledge. This is in line with the findings of Ashraf and Galor (2013). On the other hand, in line with the findings of Gomes (2017), linguistic distance acts as a barrier to information.

While this paper together with Gomes (2017) represent some of the first attempts to identify how diversity affects individual-level outcomes and the channels via which diversity might work, there remain a lot of avenues of future research. One such avenue would be establishing a causal relationship between diversity and the individual-level outcomes.

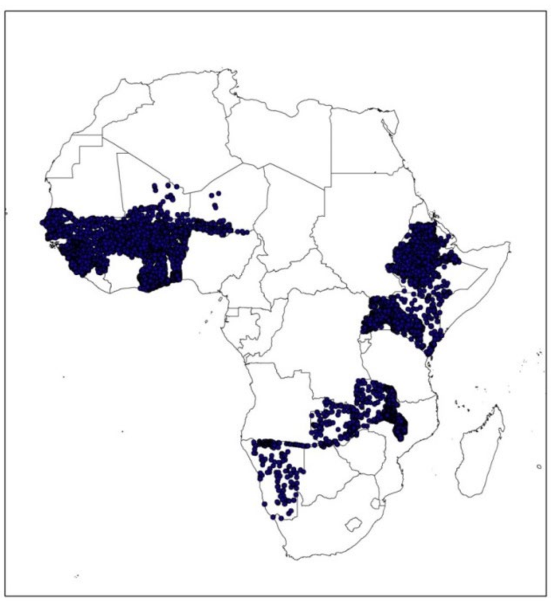
References

1. Alesina, A., R. Bakir, and W. Easterly (1999). Public goods and ethnic divisions. The Quarterly Journal of Economics, MIT Press 114 (4) November, 1243 - 1284.

[4] The results are similar for circles of larger radii. These results are available upon request.

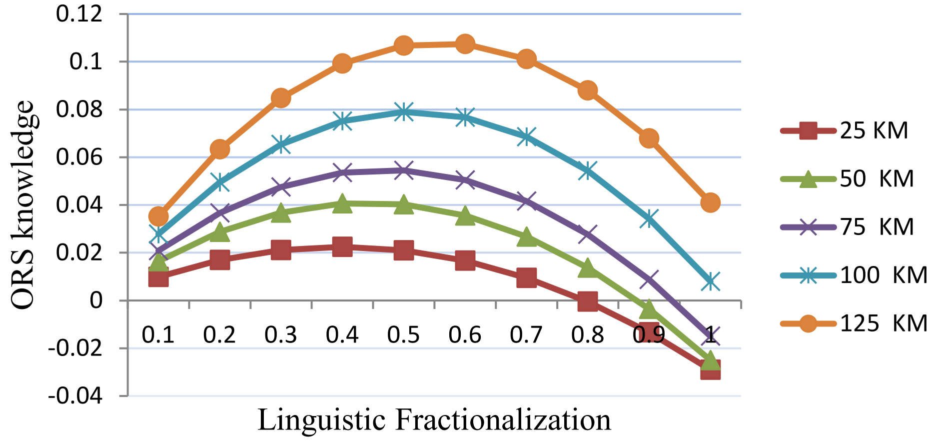
2. Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. Journal of Economic growth, 8 (2), 155-194.
3. Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. The American Economic Review, 103 (1), 1-46.
4. Desmet, K., J. F. Gomes, and I. Ortuño-Ortín (2018). The geography of linguistic diversity and the provision of public goods. NBER Working Paper No. 24694.
5. Desmet, K., Weber, S., & Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. Journal of the European Economic Association, 7(6), 1291-1318.
6. Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. Quarterly Journal of Economics 112, no.4 November, 1203-1250.
7. Glaeser, E. L., J. Scheinkman, and A. Shleifer (1995). Economic growth in a cross-section of cities. Journal of monetary economics 36 (1), 117-143.
8. Gomes, J. F. (2017). The health costs of ethnic distance: evidence from Sub-Saharan Africa. NCID Working Paper WP No. WP04/2017.
9. McLeod, P. L., S. A. Lobel, and T. H. Cox (1996). Ethnic diversity and creativity in small groups. Small group research 27 (2), 248-264.
10. Ottaviano, G. I. and G. Peri (2006). The economic value of cultural diversity: evidence from us cities. Journal of Economic Geography 6 (1), 9-44.
11. Robinson, A. L. (2013). Ethnic diversity, segregation, and ethnocentric trust in Africa. Stanford University, Unpublished Manuscript.

Figure 1: DHS Locations



Notes: DHS clusters showing the locations of the interviewed women used in the study. The study uses a sample of 205,705 women from 30 DHS surveys from 14 sub-Saharan African countries: Kenya, Uganda, Ethiopia, Burkina Faso, Malawi, Senegal, Zambia, Sierra Leone, Mali, Guinea, Ghana, Benin, Namibia and Niger. Please refer to Gomes (2017) for more details.

Figure 1: Fractionalization and the Probability of knowing about ORS



Notes: This graph demonstrates the relationship between linguistic fractionalization and the probability of knowing about the oral rehydration product (ORS). Linguistic fractionalization has been calculated in regions constructed by drawing circles of different radii around the interviewed woman. The graph above demonstrates the relationship of ORS knowledge and fractionalization for circles of 5 different radii as listed within the graph.

Table 1: Summary Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
ORS	205,705	0.76	0.43	0.00	1.00
ELF 25 km	205,705	0.41	0.27	0.00	0.92
ELF 50 km	205,705	0.49	0.26	0.00	0.92
ELF 75 km	205,705	0.54	0.25	0.00	0.94
ELF 100 km	205,705	0.58	0.23	0.00	0.93
ELF 125 km	205,705	0.61	0.22	0.00	0.94
Linguistic Distance 25 km	205,705	0.08	0.20	0.00	1.00
Linguistic Distance 50 km	205,705	0.08	0.20	0.00	1.00
Linguistic Distance 75 km	205,705	0.09	0.20	0.00	1.00
Linguistic Distance 100 km	205,705	0.09	0.20	0.00	1.00
Linguistic Distance 125 km	205,705	0.09	0.20	0.00	1.00
Wealth Index	205,705	3.01	1.42	1.00	5.00
Urban dummy	205,705	0.27	0.44	0.00	1.00
Educational Attainment	205,705	0.78	1.19	0.00	5.00

Table 2: Linguistic Fractionalization and Health Information (ORS)					
VARIABLES	25 KM	50 KM	75 KM	100 KM	125 KM
			Panel 1		
ELF	0.000295	0.00748	0.00962	0.0177	0.0278**
	(0.00835)	(0.00935)	(0.0104)	(0.0117)	(0.0128)
Observations	205,705	205,705	205,705	205,705	205,705
R-squared	0.165	0.165	0.165	0.165	0.165
			Panel 2		
ELF	0.108***	0.183***	0.229***	0.302***	0.380***
	(0.0273)	(0.0306)	(0.0371)	(0.0447)	(0.0553)
ELF squared	-0.140***	-0.213***	-0.251***	-0.303***	-0.349***
	(0.0337)	(0.0355)	(0.0402)	(0.0457)	(0.0530)
Observations	205,705	205,705	205,705	205,705	205,705
R-squared	0.165	0.166	0.166	0.166	0.166
			Panel 3		
ELF	0.113***	0.186***	0.233***	0.308***	0.386***
	(0.0273)	(0.0305)	(0.0370)	(0.0445)	(0.0549)
ELF squared	-0.142***	-0.211***	-0.248***	-0.300***	-0.345***
	(0.0337)	(0.0355)	(0.0401)	(0.0455)	(0.0527)
Linguistic Distance	-0.0375***	-0.0472***	-0.0577***	-0.0680***	-0.0756***
	(0.0113)	(0.0120)	(0.0124)	(0.0134)	(0.0144)
Observations	205,705	205,705	205,705	205,705	205,705
R-squared	0.165	0.166	0.166	0.166	0.166

Notes: Robust standard errors clustered at the DHS cluster level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in each specification is a 0-1 binary variable indicating whether the interviewed woman has either heard of or used the Oral Rehydration product (ORS) for treating children with diarrhoea. The relevant regions in which the ELF and the linguistic distance variables are calculated are circles of different radii drawn around the mother. These radii are listed in the column headings. All the specifications control for ethnicity, religion, country and birth-cohort fixed effects; log population in the circle; a dummy for urban areas; dummies for educational attainment of the woman and the log of distance to the capital.