



## **A Pluri-Ethno-Linguistic Fragmentation Index**

Stéphanie Cassilde, Kelly Labart

### **► To cite this version:**

Stéphanie Cassilde, Kelly Labart. A Pluri-Ethno-Linguistic Fragmentation Index. *Revista Internacional de Organizaciones = International Journal of Organizations*, 2019, 23, pp.223-242. <halshs-02909924>

**HAL Id: halshs-02909924**

**<https://shs.hal.science/halshs-02909924v1>**

Submitted on 31 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# A Pluri-Ethno-Linguistic Fragmentation Index<sup>1</sup>

Stéphanie Cassilde

Centre d'Etudes en Habitat Durable (CEHD), Belgium

stephanie.cassilde@cehd.be

Kelly Labart

Foundation for Studies and Research on International Development (FERDI)

kelly.labart@ferdi.fr

*Abstract: The Ethno-Linguistic Fragmentation index aims to measure the probability of two random individuals belonging to the same group. However, a person may belong to several groups; therefore, the Pluri-Ethno-Linguistic Fragmentation index generalizes the ELF index to consider multiple belongings within a country. PELF index results are discussed in relation to the crucial case of Luxembourg, which provides evidence of the validity of our new proposed index. The PELF index deepens our understanding of the repercussions of the coexistence of several languages in a society.*

*Keywords: ELF index, fragmentation, Luxembourg, multiple belonging, PELF index, plurilingualism, fractionalization*

---

<sup>1</sup> This research into PELF index started in 2012 when Stéphanie was working in Luxembourg at the CEPS/INSTEAD (Centre d'Études de Populations, de Pauvreté et de Politiques Socio-Economiques/International Networks for Studies in Technology, Environment, Alternatives, Development, i.e. the current LISER). It was presented at Centre d'Etudes et de Recherches sur le Développement International (CERDI, France) research seminar, 2013, and at the "Language at Work. Research Advances in Social Sciences" International Conference, 2018, Tarragona, Catalonia, Spain. We warmly thank the participants for their comments and feedback, and also the anonymous referees for the salient points underlined. This research has also benefited from early discussions with our colleagues Marie Valentova, Vianney Dequiedt, Michel Tenikue, Simone Bertoli, and Adeline Gilson. We warmly thank them.

## Introduction

Diversity of languages, multilingualism and plurilingualism are seen as valuable assets for culture, social cohesion and development. One of the many ways to describe language and ethnic diversity, the ethno-linguistic fragmentation index (ELF index) was built by Taylor and Hudson (1972) on the basis of the ethnic data collected in the 1960s by Soviet researchers in *The Atlas Narodov Mira*. Other ways have also been developed to take into account ethno-linguistic fragmentation and how it affects political systems (Almond, 1956), for example, democracy (Reilly, 2001) (see Ginsburg and Weber (2016) for a detailed approach toward relations between economics and language for instance). However, we focus solely on the ELF index because it has crystallized a huge part of the debate.

The ELF index calculates the probability that two individuals randomly drawn from a population are from different ethno-linguistic groups. It aims to observe the coexistence of several languages in a society. However, the index does not take into account multiple ethno-linguistic backgrounds. The ethno-linguistic groups are seen as separate and exclusive from each other: “An ethnic group or an ethno-linguistic group is a collection of people who share the same language or have a common culture based on language” (Lane and Ersson, 1998: 54). Consequently, each society (country) is considered to be divided into exclusive groups that may be described in terms of share of the population. In terms of languages, the ELF index focuses on multilingualism and ignores plurilingualism. Multilingualism may be defined in numerous ways (Coulmas, 2018), and some of these definitions may blur the line between multilingualism and plurilingualism. Following Trim (2007) we define multilingualism as the coexistence of languages among the population and plurilingualism as each individuals’ use of languages. This distinction between multilingualism and plurilingualism has huge implications regarding our understanding of social cohesion and other social and economic aspects. Growing migrations and increasing diversity within societies lead to multiple belongings in terms of languages and to pluriethnicity, where an individual belongs to several ethnic groups, exists (see Root (1992 and 1996)). Multiple belonging may also be recognized in official statistical systems. For instance, since 2000, respondents in the United States of America may tick several options to declare their race. In terms of plurilingualism, historical perspectives remind us that it is “one of the paradigms of the socio-cultural and socio-institutional history of Europe, as it is in other parts of the Old World” (Grévin, 2016: 339; our translation). Current migration streams may further increase cases of multiple belonging.

Current findings linking ethno-linguistic fragmentation to various socio-economic phenomena are based on this initial focus on multilingualism. Whereas diversity has been recorded as a positive an engine for economic and social development at the micro level, at the macro level it has also been mentioned as a negative factor (Klein, 2003; Williams, 2006; Wanyonyi, 2012). For instance, the ELF index was used as an explanatory variable for low quality institutions, including democracy (Rabushka and Shepsle, 1972; Hadenius, 1992; Mauro, 1995; Alesina *et al.*, 2003), bad spending on public goods and provisions (Alesina *et al.*, 1999; Habyrimana *et al.*, 2007), poor social participation (Alesina and La Ferrara, 2000a; Vigdor, 2004), poor trust (Alesina and La Ferrara, 2000b), poor political stability (Annett, 2001), low quality of life (Nettle, 2000), low quality of government (La Porta *et al.*, 1999), poor economic performance (Easterly and Levine, 1997; Rodrik, 1998; Temple, 1998; Easterly, 2001; Alesina *et al.*, 2003; Posner, 2004; Alesina and La Ferrara, 2005; Campos and Kuzeyev, 2007; Fedderke *et al.*, 2008), and the presence of conflicts (Collier *et al.*, 2001; Easterly, 2001; Collier and Hoeffler, 2002). All of these studies assert that multilingualism (and multiethnicity), i.e. the coexistence of a given number of separate and exclusive linguistic (and ethnic) groups, is harmful for socio-economic outcomes. The consequences of this finding in terms of public policy recommendations are considerable because the logical response is to foster the least diverse society possible in order to ensure better outcomes. This finding also conflicts with the positive outcomes that may be observed on almost a daily basis. Furthermore, ongoing societal and economic changes seem to be continuing in the direction of greater diversity. Another consequence of this finding relates to the linguistically diverse continent of Africa. The study by Easterly and Levine (1997), *Africa's Growth Tragedy*, has led to widespread criticism, notably from Arcand *et al.* (2000), *How to Make a Tragedy*, who underlined the empirical problems of falling into the temptation of identifying linguistic diversity as a key explanation for Africa's problems. Other studies that are critical of the manner in which the ELF index has been used are Reynal-Querol and Montalvo (2000), Reynal-Querol (2002), and Montalvo and Reynal-Querol (2005) (see Labart (2010) for a survey). Some research has focused on criticism about the way ethnic groups are defined in the ELF index (Fearon, 2003; Bossert *et al.*, 2011). Others, in contrast, have described the index as a positive explanatory variable regarding democracy (see Reilly (2000-01) and the crucial case of Papua New Guinea), civil wars (Fearon and Laitin, 2003) and conflict levels (see Gardeazabal (2011) and the Basque country). They also point to the importance of understanding the ELF index as a means for measuring potential social and political tensions between groups. Some

research has shown that the relation between linguistic fragmentation and other macro indicators may depend on a threshold. Desmet et al. (2012) focused on the depth of linguistic cleavages: deep linguistic cleavages have a significant effect on civil conflicts and redistribution, whereas more superficial linguistic cleavages have a greater effect on economic growth and public goods. Regarding national identification, Masella (2011) found no evidence. The research about the socio-economic effects of ELF on multilingualism and multiethnicity, as measured by the ELF index, are still ongoing. Straveren and Pervaiz (2017) challenged the assumed link between the ELF index and social cohesion, while Alesina et al. (2016) showed how the correlation with GDP per capita vanishes when control for ethnic group inequality is included. Recently, Arcand and Jaimovich (2019) concluded that “ethnic heterogeneity is unlikely to be a driver for sub-optimal economic exchanges”.

As part of the debate, other variants of the ELF index have led to seminal generalization formulae, which are described below. However, the existing generalizations still focus on multilingualism and multiethnicity; that is, like the ELF index they are Multi-Ethno-Linguistic indexes. The objective of this article is to offer an index which shifts from multilingualism to plurilingualism – and from multiethnicity to pluriethnicity – to enable further discussion of the relation between fragmentation and other phenomena. Until now, the ELF index and its generalizations were calculated with groups that are defined at a macro level and as strictly separately from each other. However, in real-life multi-ethnic and multilingual societies, a person might belong to several macro groups simultaneously by being multi-racial, binational and / or plurilingual. Moreover, multiple linguistic belonging is the norm throughout the world (Aronin and Singleton, 2010). The present article aims to tackle this issue and proposes the construction of a pluri-ethno-linguistic fragmentation index (PELF index) to take into account membership of several ethnic or linguistic groups at the same time. This shift toward plurilingualism and pluriethnicity also means that empirical research on the PELF index must be based on micro data rather than macro measurements of population share, which is a point of similarity with the generalized (multi) ethno-linguistic fractionalization index of Bossert et al. (2011). Micro data were also used by Caselli and Coleman (2006) regarding ethnic distance and linguistic distance.

The PELF index is a generalization of the ELF index and its main advantage is that it provides a measurement of ethno-linguistic fragmentation that takes into account the real communication practices of individuals within a country; it can also be used to determine the real multiple ethnic belonging of individuals. PELF

significantly enhances the focus on languages, in contrast to previous studies in which linguistic information has often been interpreted as simply as a proxy for ethnicity or culture. The L of the PELF index matters per se. The implications of the PELF index are discussed in relation to the crucial case of Luxembourg, which has been chose because it directly contradicts the assertion that multilingualism (and multiethnicity) have a harmful effect on socio-economic outcomes. Indeed, whereas several African countries have been highlighted because they combine poor socioeconomic performance with a high score on the ELF index (Easterly and Levine, 1997), Luxembourg also has a high score on the ELF index but one of the top socioeconomic rankings in the world.

Section 1 describes the establishment of the PELF index as a generalization of previous indexes, notably of the ELF index. Section 2 explains the reasons why Luxembourg was chosen for the case study and gives a description of the data. The results are presented and discussed in Section 3.

### 1. From the ELF index to a PELF index

The initial ELF index is an application of the Herfindahl formulae to ethno-linguistic groups (Taylor and Hudson, 1972):

$$ELF = 1 - \sum_{i=1}^I \left(\frac{n_i}{N}\right)^2 \in [0,1] \text{ and } \sum_{i=1}^I n_i = N$$

where  $n_i$  is the population of group  $i$  within the total population  $N$ .  $I$  is the number of groups. The ELF index is the probability that two individuals randomly drawn in a population are from different ethno-linguistic groups. The data are based on dividing the population into languages/ethnicities, and one individual can belong solely to one language/ethnicity. Each group speaks one single language. Each group has one single ethnicity. The higher the ELF index, the more fragmented the population from a multilingual/multiethnic point of view.

Fearon (2003) showed that groups strictly separated from each other might be more or less culturally close. On the basis of linguistic distance, he defined a continuous resemblance factor  $r_{ij} \in [0, 1]$ , where 0 means that two groups' languages,  $i$  and  $j$ , come from completely different families, and 1 means that the two groups speak the same languages. The ELF index becomes:

$$Cult frag = 1 - \sum_{i=1}^I \sum_{j=1}^I \frac{n_i}{N} \frac{n_j}{N} r_{ij} \in [0,1] \text{ and } \sum_{i=1}^I n_i = N \text{ and } \sum_{j=1}^I n_j = N$$

with the same notation as above. This formula is still established from a multilingual/multiethnic point of view.

In the light of Okediji's (2005) insight that ethnic diversity needs more than language to be described, Bossert *et al.* (2011) offered a Generalized Ethno-Linguistic Fractionalization index (GELF):

$$G(S) = 1 - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n S_{ij}$$

with the same notation as above and with  $S$  as the similarity matrix regarding individuals'  $i$  and  $j$  characteristics. It measures the expected dissimilarity between two individuals drawn at random from a society. This socially informed measure of ethnic diversity uses similarities (i.e. several characteristics collected in micro data) to build groups and emphasise ethnicity matters in certain contexts but not in others.

Regarding the newly proposed PELF index, we shift from social multilingualism to individual multilingualism in accordance with the distinction put forward by Baka (2005). Individual multilingualism corresponds to what Trim (2007) labels as plurilingualism, hence the name of the PELF index. Groups are still separated from each other, but they are plurilingual rather than unilingual. Given that plurilingualism implies the need of micro data rather than macro data,  $K$  is the number of groups, defined at a micro level, i.e. taking into account the linguistic skills of each individual rather than the aggregated proportion for each language spoken within the country. Individuals are not random, as the computation relies on their linguistic uses. The  $K$  groups are predefined on the basis of linguistic uses regarding the condition of a certain society. The PELF index is:

$$PELF = 1 - \sum_{i=1}^K \sum_{j=1}^K \frac{n_i}{N} \frac{n_j}{N} o_{ij} \in [0,1] \text{ and } \sum_{i=1}^K n_i = N \text{ and } \sum_{j=1}^K n_j = N$$

with the same notation as above.

An overlap factor  $o_{ij}$  takes the value of 0 if groups  $i$  and  $j$  do not have a joint language, and the value of 1 otherwise. When  $o_{ij}$  takes the value of 0 systematically, i.e. when each individual is unilingual,  $PELF = ELF$ .

The overlap factor  $o_{ij}$  may be also seen as being similar to the resemblance factor used by Fearon (2003) in that it takes into account the structural relationship between languages and can thus uses cultural similarities to temper the score for ethnic/linguistic fractionalization. At a micro level, the choice of the possible values of  $o_{ij}$  relies on how plurilingualism is measured. If linguistic practices are chosen (e.g. is a given language spoken or not in a given context) then these are given a binary value of 0 or 1, whereas if linguistic skills are



chosen, then the values have to be on a continuous scale to take into account the heterogeneity of skills. For example, an individual with a level of A1 in a L2 has, in principle, lower plurilingual capabilities than someone who has level C2 in the same L2. Languages matter in the PELF index, they are not a proxy for ethnicity. PELF supports the investigation of multiple belongings to languages (PELF), and ethnicity (PELF).

As for the ELF index, the higher the value of the PELF index, the higher the fragmentation. The PELF index is equal to zero when there is no fragmentation at all: either the whole society is unilingual, or all the individuals speak the same set of languages. The value of the PELF index is expected to be smaller than the value of the ELF index. Thus, some highly fragmented countries from the point of view of the ELF index might no longer be fragmented from the point of view of the PELF index. This might change the correlation between fragmentation and socioeconomic phenomena.

The contribution of PELF index is three fold. First, it includes groups that speak more than one language, whereas previous indexes do not. Second, it relies on real linguistic practices, beyond resemblance between languages. Third, it treats language in its own right and not as a proxy for other variables.

In terms of development policies, the PELF index may produce difficult findings for some African countries. Many of them simultaneously score low on socioeconomic indicators and high on ethnolinguistic fragmentation (Easterly and Levine, 1997), while the extent of plurilingualism in these countries is important.

The magnitude of the difference between the PELF and ELF indexes cannot be set a priori, which invites us to search for the meaning for this difference. This will be discussed in the results section.

**2. Application:** The Case of Luxembourg for groups that speak up to two languages in three contexts

*a) The crucial case of Luxembourg: simultaneity of high fragmentation and high GPD per capita*

The crucial case of Luxembourg has been chosen because it offers a direct contradiction of the assertion that a plurilingual environment is a contributing factor to low socioeconomic performance. The Grand-Duchy of Luxembourg has three institutional languages: Luxemburgish (See Horner and Weber, 2010), French, and German. Civil servants are expected to be plurilingual, i.e. to master these three languages. The schooling system promotes plurilingualism too in that infant education uses Luxemburgish, with German and French introduced in the first and second years respectively of national primary school. Moreover,



several administrative documents are provided in Portuguese and English, thus reflecting the diverse population of Luxembourg. In terms of the former, a significant share of foreigners living there comes from Portugal and several Luxemburgish nationals have Portuguese ancestry. Regarding English, EU employees have been a significant immigration inflow in the country since the 1980s. Using English enables them to engage with administrative duties during their stay in the country.

The share of foreigners among Luxembourg inhabitants is high (43.8%, STATEC), with 38% of its inhabitants born abroad (OCDE, 2012). This figure cannot be used to approximate the share of Luxemburgish nationals with foreign ancestry as one can be born in Luxembourg with foreign parents and then choose Luxemburgish nationality when one becomes an adult.

Table 1: Description of Luxembourg and of 12 ACP African Countries\*

	area (sq.m.)	inhabitants**	access to the sea	GDP per capita rank PPP (World Bank)	national language(s)***	institutional languages****	overall number of living languages****
Luxembourg	2586	524900 (2011)	no	1	2	3	3
3 ACP African Countries with the highest ELF index (Easterly and Levine, 1997)							
Cameroon	475442	20129878 (2012)	yes	140	2	11	280
Congo (DRC)	2345000	4366266 (2012)	no	181	1	8	212
Tanzania	945203	45798475 (2012)	yes	158	1	3	126
3 ACP African Countries with increasing fragmentation (Easterly and Levine, 1997; Alesina et al., 2003)							
Benin	112622	9598787 (2012)	yes	156	1	8	54
Burundi	27834	84320987 (2012)	no	178	1	3	3
Ethiopia	1104300	73750932 (2007)	no	170	1	30	87
3 ACP African Countries with a decreasing fragmentation (Easterly and Levine, 1997; Alesina et al., 2003)							
Equatorial Guinea	28050	616459 (2011)	yes	21	2	4	14
Swaziland	17363	1403362 (2013)	no	102	2	4	5
Zimbabwe	390245	12619600 (2012)	no	186****	1	10	20
3 ACP African Countries with the smallest area							
Cape Verde	4033	426998 (2011)	island	122	1	2	2
Gambia, The	11300	1735464 (2011)	yes	150	1	4	12
Seychelles	455	82247 (2011)	island	37	3	3	3

Nota Bene:

\*: The African, Caribbean, and Pacific Group of States has specific EU-partnership conditions in order to foster the development of its members, who face higher difficulties regarding their integration into world trade. This is relevant because it put the case of Luxembourg into perspective. Of this group of states, 12 African countries were chosen because (i) they belong to Easterly and Levine's (1997) highlighted region where fragmentation has a negative effect on growth, (ii) they show variability in other relevant areas (highest ELF index, increasing or decreasing fragmentation, and smallest areas).

\*\* : Figures from STATEC (Luxembourg), CIA World Factbook (Benin, Burundi, Cameroon, Cape Verde, Congo (DRC), Equatorial Guinea, Gambia, Seychelles, Swaziland, Zimbabwe), National Bureau of Statistics (Tanzania), Central Statistical Agency (Ethiopia).

\*\*\*: From Lewis et alii (2003) from SIL International ([www.ethnologue.com](http://www.ethnologue.com)). The national language "is used in education, work, mass media, and government at the national level", the institutional languages are those used by at least one institution, living languages are those from the country and currently used within the country. It seems that languages of migrants are not taken into account. Indeed, many more languages were declared for Luxembourg in EPC 2009.

\*\*\*\*: From the International Monetary Fund.

Regarding institutional multilingualism, the situation is similar for several African countries (cf. Table 1). Fardon and Furniss (1994: i) declared that “multilingualism is Africa’s lingua franca”. It is also true at individual level through plurilingualism (Badejo, 1989; Trudell, 2009). Diki-Kidiri speaks about the “multilinguisme convivial d’Etat” (State friendly multilingualism) to describe several African countries: « The concept of conviviality is based on a vision of relations between languages that is philosophically optimistic, socially serene, and economically profitable »<sup>2</sup> (2004: 27). Conviviality is an accurate description of what is practiced in Luxembourg.

However, the difference in GDP purchasing power parity per capita between these countries and Luxembourg is huge (cf. Table 1). Therefore, Luxembourg is crucial regarding the link between the economic development of a country and its ethnic or linguistic fragmentation. Indeed, while Luxembourg has been ranked first in terms of GDP per capita for years, its ELF index is as high as that of several African countries (Data available for all countries on the Foundation for Studies and Research on International Development website: [www.ferdi.fr](http://www.ferdi.fr)). Except for the first calculation of its index on the basis of *The Atlas Narodov Mira*, which shows the share of Luxembourgish, French and German speakers only, it is among the highest values. Moreover, some African countries have a lower index than Luxembourg for ethno-linguistic, ethnic, and linguistic fragmentations (Burundi), or for ethnic and linguistic fragmentations (Equatorial Guinea, Cape Verde, Seychelles, Swaziland, and Zimbabwe).

*b) The data: two main languages used in three daily environments*

The Cultural Practices Survey 2009 data set (Enquête sur les Pratiques Culturelles 2009 / EPC 2009) was collected by CEPS/INSTEAD for the Luxembourgish Department of Culture in order to shed light on contemporary cultural practices in Luxembourg. Face-to-face questionnaires were conducted with individuals chosen through a non-proportional random stratified sampling procedure. The Stata survey procedure is used and includes the appropriate weight. The sample of 1880 observations is representative of the whole resident population of Luxembourg except for European civil servants. This delimitation of the sample is in accordance with the primary focus of the survey, which is to monitor cultural practices. Regarding linguistic practices, EU employees may influence daily uses. Also, about half of the labor force enters and exits the country from France, Belgium, and Germany each working day, which may influence linguistic practices at work too.

<sup>2</sup> « Le concept de convivialité se fonde sur une vision des rapports entre les langues qui se veut philosophiquement optimiste, socialement sereine, et économiquement rentable »

Nationality and mother tongue variables allow ELF index to be computed in order to provide a benchmark regarding other fragmentation data sets and regarding the PELF index. Only one nationality per person is collected in the EPC 2009, even though since 2008 foreigners have been able to acquire Luxembourgish nationality without renouncing their birth nationality. Nationality is used to compute Alesina *et al.*'s (2003) ethnic fragmentation index, and mother tongue is used to compute their language fragmentation index.

The PELF index is computed for three different environments. Indeed, the advantage of EPC 2009 is that interviewees were asked to indicate which was the first and the second language they used at work (or at university), with friends, and at home (with the family). Interviewees were first offered a list of six languages (Luxembourgish, French, German, Portuguese, English, and Italian), and then a list of countries if the appropriate language was not among the initial list. This feature of the data collection also means that the PELF can overestimate the level of linguistic fragmentation as several languages may be used in a given country. Although the PELF index formulae take into account plurilingualism whatever the number of languages spoken by individuals, here two main languages are collected and thus used for the calculation. The drawback of having solely three environments is that only a part of the plurilingualism in Luxembourg is taken into account because the other languages that the interviewees might use in life situations that were not listed in the questionnaire are not covered. Thus, the PELF index with EPC 2009 should be considered as an overestimation of fragmentation. Language fragmentation indexes are also calculated for these three environments as benchmark.

To calculate the PELF index, the distribution of the  $K$  linguistic groups is first calculated. For those who do not know which language they are using, a “do not know” linguistic group is included; this group is not considered as being able to communicate with the other groups, including with itself. The results with and without the “do not know” group change at the third decimal of the PELF index values.

The overlap factor  $\sigma_{ij}$  takes here the value of either 0 or 1, as we observe linguistic practices and not potential linguistic interactions.

Regarding the working / academic environment, only active residents (including students) are considered, i.e. 1401 observations. On the basis of the most used first language in this context, there are 16 linguistic groups and one “do not know” group ( $l = 17$ ): the main group is for Luxembourgish (47.12%), then for French (36.17%). Including the most used second language, plurilingualism is high: 82.31% speak at least two languages; 0.59% do not know which language

they most speak in this context. There are five unilingual groups and 40 bilingual groups ( $K = 46$ ): the main group is now bilingual (Luxembourgish and French, 44.52%; and then French and Portuguese, 11.16%).

There are 1879 observations for the friendship environment. One individual declared that they had no friends. On the basis of the most used first language in this context, there are 35 linguistic groups and one “do not know” group ( $l = 36$ ): the main group is for Luxembourgish (60.33%), then for French (20.03%). Including the most used second language, plurilingualism is also high: 79.06% speak at least two languages; 0.07% do not know which language they most speak in this context. There are 13 unilingual groups and 78 bilingual groups ( $K = 92$ ): the main group is now bilingual (Luxembourgish and French, 37.36%; and then French and Portuguese, 12.97%).

There are 1874 observations for family environment. Six individuals declared that they had no family. On the basis of the most used first language in this context, there are 48 linguistic groups and one “do not know” group ( $l = 49$ ): the main group is for Luxembourgish (55.90%), then for Portuguese (16.30%). Including the most used second language, plurilingualism is lower than for the other environment: 50.49% speak at least two languages; 0.05% do not know which language they most speak in this context. There are 29 unilingual groups and 102 bilingual groups ( $K = 132$ ): the main group is still Luxembourgish (33.82%) and the second group is bilingual (Luxembourgish and French, 15.45%).

Beyond the two main groups without and with plurilingualism, the distributions of the  $K$  linguistic groups in these three contexts are consistent with the linguistic groups distributions of Dickes and Berzosa (2010a), and with their language relationships analysis. Notably, French is the main mediatory language (Dickes and Berzosa, 2010a and 2010b) and Portuguese, the language of the main group of foreigners, is a linguistic pillar (Dickes and Berzosa, 2010a).

### 3. Results and Discussion

Ethnic and language fragmentation index are extremely close between 2003 and 2009 (*cf.* Table 2). The values are slightly higher in 2009 than in 2003, which can be explained by an increased number of foreigners among the inhabitants of Luxembourg (this share was 29.4% in 1991, 36.9% in 2001, 38.1% in 2003, and 43.7% in 2009; STATEC). The index by Alesina et al. (2003) is based on data collected from 1997 to 2001. There were no figures for foreigners in STATEC’s state of population information for 1997. The values of the language fragmentation index for each of the three linguistic environments is even closer

to Alesina *et al.*'s value (2003). The degree of language fragmentation in a working context and at home are similar, while it is slightly less important with friends. As an additional benchmark we can also turn to the Linguistic Diversity Index (LDI), which is the share of the world's languages in a country divided by this country's percentage of the world population. Luxembourg also scores highly on the LDI (LDI of Luxembourg is 10; see Coulmas (2018) to situate Luxembourg regarding other countries).

Table 2: Ethno-Linguistic and Pluri-Ethno-Linguistic Fragmentation Index for Luxembourg

	ELF <sub>e</sub> Ethnic Fragmentation index		ELF <sub>l</sub> Language Fragmentation index		PELF index
nationality	0.53*	0.59	-	-	-
mother tongue	-	-	0.644*	0.68	-
at work / the university	-	-	0.64	-	0.18
with friends	-	-	0.58	-	0.14
at home (with the family)	-	-	0.64	-	0.43

Data: Except for \*, where Alesina *et alii* (2003) values are indicated, all values are calculated on the basis of EPC 2009 (CEPS/INSTEAD / Luxembourgish Department of Culture).

Including plurilingualism significantly decreases the value of the fragmentation index (*cf.* Table 2). As expected, the PELF index is lower than the ELF index. From this perspective, Luxembourg is not fragmented, especially in a working or a friendship environment (respectively, PELF = 0.18 and PELF = 0.14). With 92 and 132 linguistic groups (defined as *K*), plurilingualism leads to the ELF index value of 1964 calculated with three ethno-linguistic groups only. Fragmentation in a familial environment decreases, falling from 0.64 to 0.43 when plurilingualism is taken into account.

Going back to the definition of multilingualism and plurilingualism in the introduction, the difference between the ELF and PELF indexes offers interesting interpretations. Indeed, the high fragmentation in terms of the coexistence of languages among the population (multilingualism) compared with the low fragmentation from the perspective of individuals' uses of languages (plurilingualism) shows a kind of linguistic cohesion regarding language diversity, *i.e.* the extent of the linguistic overlaps highlighted by plurilingualism. The results for Luxembourg suggest that there is a high level linguistic cohesion at work and with friends while the familial environment is kept more separate and thus more diverse.

On the one hand, the lower value of fragmentation found when using the PELF restores the potential link between high economic outputs and low

linguistic fragmentation. On the other hand, the contribution of the PELF shows that it needs to be applied to all countries in order to enrich research through macro and micro measurements of languages and to shed light on the influence of the PELF on various countries' indicators.

In the current study (due to the questionnaire design), the definition of plurilingualism is limited to the two main languages used at work, at home and with friends. This represents a limitation to our findings given that the value for all the languages used by individuals is expected to lower the value of the PELF index again, i.e. to lower the fragmentation. Also, plurilingualism here is also based on self-declarations of use and does not take into account linguistic skills themselves, i.e. potential plurilingualism. The expected effect of this shift in interpreting the PELF interpretation is less straightforward. Indeed, knowledge of a language can be heterogeneous, and even being proficient neither implies its use a willingness to communicate with the persons who use this language. For instance, Schooland et al.'s (2010) social language network analysis is based on effective communication rather than on linguistic skills. For example, they considered the following languages to be separate when calculating the PELF index because interviewees clearly mentioned one or another: Bosnian, Serbian, Croatian, Albanian / Serbian (Kosovo), and Serbo-Croatian (Yugoslavia). As a lower PELF index value based on the aggregation of these languages could not be interpreted a priori as a lower fragmentation level, these languages were separated to consider the reality of the linguistic exchanges: languages go beyond being a skill. Finally, one may be careful about which set of languages is offered to respondents: indeed, the list is often limited to written languages while unwritten languages are omitted; This may be a crucial point regarding certain linguistic areas.<sup>3</sup>

The novelty of the PELF index is that it takes plurilingualism into account. However, it does so by assuming plurilingualism to be unique while regarding plurilingualism and multilingualism as heterogeneous and dynamic (Juillard, 2005; Djité, 2009; Coulmas, 2018). Any interpretation of the PELF index per se and of its effects, when used as an explanatory variable, should consider the linguistic changes over space and time. Like the ELF (Luiz, 2015), the PELF is not constant over time, which raises specific concerns regarding econometric considerations. Moreover, it is one thing to explain plurilingualism and multilingualism, it is quite another thing to explain fragmentation, i.e. the relationships between ethnic or linguistic groups (whether defined as *l* or *K*; See Alesina and La Ferrara (2000a and 2005) and Leeson (2005) regarding the debate about ethno-linguistic fragmentation endogeneity).

<sup>3</sup> We warmly thank Florian Coulmas for bringing this point to our attention.

## 4. Conclusion

This article proposes a Pluri-Ethno-Linguistic Fragmentation index (PELF index) in order to consider the real uses of languages in multi-ethnic and multilingual societies. The PELF index is a theoretical continuation of the existing ELF index; that is, it shifts from multilingualism to plurilingualism (from social multilingualism to individual multilingualism) when describing the ethno-linguistic structure of a country where a person might speak several languages at the same time.

The application of the PELF to the crucial case of Luxembourg provides evidence of the validity of our new proposed index. It supports the theoretical expectation: the value of the PELF index is smaller than the value of the ELF index. Based on the first two most used languages at work, at home, and with friends, the PELF in this context still overestimates linguistic fragmentation.

Given that plurilingualism decreases the linguistic fragmentation of Luxembourg from 0.64 to 0.18 in a working environment, from 0.58 to 0.14 in a context of friendship, and from 0.64 to 0.43 in a familial context, the PELF index shows that the fragmentation is significantly lower when tangible linguistic uses are taken into account. The difference between the ELF and PELF indexes for Luxembourg suggests that there is a high level of linguistic cohesion at work and with friends. The familial environment is more linguistically separated.

This article deepens understanding of the social repercussions of the coexistence of several languages in a society. Its findings do not make a connection between fragmentation and institutional efficiency, democracy, trust, political stability, public good spending, social participation, the quality of life or institutions, conflicts, economic performance or national identification. Case studies of countries where the relevant micro data sets with variables giving insights about plurilingualism are available should shed light on these phenomena. They might also compare the effects of a PELF index calculated on the basis of language uses with those of a PELF index based on linguistic skills. Also, the reverse effect on plurilingualism of these economic, social, and political phenomena still needs to be addressed (Bangura, 2010).

Luxembourg is a crucial case because it has high number of languages, a high ELF index, high socioeconomic indicators, and PELF index values which are lower than ELF values. The future developments of this research rely first on identifying the relevant micro data sets to calculate the PELF for other countries, in particular African countries. Second, the PELF index will be used to test relations with the above mentioned macro indicators. The distance between



the ELF and PELF indexes should also be researched to investigate the social cohesion capability within a society.

The E of PELF index was not addressed because the appropriate data were not available in the EPC 2009; consequently, there is still room for improving on the present study and, therefore, for further research. Also, several aspects of plurilingualism are closely related to those of mixing between ethnic groups and how the borders between them are variously conceived (Gadet and Varro, 2006; Guyot, 2007). Given that contemporary societal changes are characterized by increases in migrations and diversity, these factors should contribute to the analysis of the interaction of diversity and fragmentation.

## 5. References

- ALESINA, Alberto; BAQIR, Reza and EASTERLY, William (1999) "Public Goods and Ethnic Divisions", *The Quarterly Journal of Economics* 114 (4): 1243-1284.
- ALESINA, Alberto; DEVLEESCHAUWER, Arnaud; EASTERLY, William; KURLAT, Sergio and WACZIARG, Romain (2003) "Fractionalization", *Journal of Economic Growth* 8: 155-194.
- ALESINA, Alberto and LA FERRARA, Eliana (2000a) "Participation in Heterogeneous Communities", *The Quarterly Journal of Economics* 115 (3): 847-904.
- ALESINA, Alberto and LA FERRARA, Eliana (2000b) "The Determinants of Trust", *NBER Working Paper Series*, Working Paper 7621.
- ALESINA, Albert and LA FERRARA, Eliana (2005) "Ethnic Diversity and Economic Performance", *Journal of Economic Literature* 43 (3): 762-800.
- ALESINA, Albert; STELIOS, Michalopoulos and PAPAIOANNOU, Elias (2016) "Ethnic Inequality", *Journal of Political Economy* 124(2): 428-488.
- ALMOND, Gabriel A. (1956) "Comparative Political Systems", *Journal of Politics* 18 (3): 391-409.
- ANNETT, Anthony (2001) "Social Fractionalization, Political Instability, and the Size of the Government", *IMF Staff Papers* 46: 561-592.
- ARCAND, Jean-Louis; GUILLAUMONT, Patrick and GUILLAUMONT-JEANNENEY, Sylviane (2000) "How to make a tragedy: on the Alleged Effect of Ethnicity on Growth", *Journal of International Development* 12 (7): 925-238.
- ARCAND, Jean-Louis and JAIMOVICH, Dany (2019) "Does ethnic diversity decrease economic interactions? Evidence from exchange networks in rural Gambia", *Economics of Transition and Institutional Change* 27(2): 327-353.
- ARONIN, Larissa and SINGLETON, David (2010) "Introduction. The Diversity of Multilingualism", *International Journal of the Sociology of Language* 205: 1-5.

- BADEJO, B. Rotimi (1989) "Multilingualism in Sub-Saharan Africa", *Africa Media Review* 3 (2): 40-53.
- BAKA, Jean (2005) "Problématique pour une meilleure gestion du multilinguisme social de type afro-européen en Afrique au Sud du Sahara", *Africa Focus* 18(1-2): 95-118.
- BANGURA, Abdul Karim (2010) "Measurable effects of multilingualism in Africa", *International Journal of the Sociology of Language* 146: 111-117.
- BOSSERT, Walter; D'AMBROSIO, Conchita and LA FERRARA, Eliana (2011) "A Generalized Index of Fractionalization", *Economica* 78 (112): 723-750.
- CAMPOS, Nauro F. and KUZEYEV, Vitaliy S. (2007) "On the Dynamics of Ethnic Fractionalization", *American Journal of Political Science* 51 (3): 620-639.
- CASELLI, Francesco and WILBUR, John Coleman (2006) "On the Theory of Ethnic Conflict", *NBER Working Paper* n°12125.
- COLLIER, Paul and HOFFLER, Anke (2002) "Greed and Greavance in Civil War", *CSAE Working Paper* 2002-01.
- COLLIER, Paul; HONOHAN, Patrick and MOENE, Karl Ove (2001) "Implications of Ethnic Diversity", *Economic Policy* 16(32): 129-166.
- COULMAS, Florian (2018). *An introduction to multilingualism. Language in a changing world*. Oxford: Oxford University Press.
- DESMET, Klaus; ORTUÑO-ORTÍN, Ignacio and WACZIARD, Romain (2012) "The political economy of linguistic cleavages", *Journal of Development Economics* 97: 322-338.
- DICKES, Paul and BERZOSA, Guayarina (2010a) "Pays multiculturel, pays multilingue? Un modèle pragmatique pour l'analyse des relations langagières au Luxembourg", *Les Cahiers du CEPS/INSTEAD / Population & Emploi* n°2010-16.
- DICKES, Paul and BERZOSA, Guayarina (2010b) «Les compétences linguistiques auto-attribuées» *Les Cahiers du CEPS/INSTEAD / Population & Emploi* n°2010-19.
- DIKI-KIDIRI, Marcel (2004) "Multilinguisme et politique linguistique en Afrique", *Actes du colloque international "Développement durable: leçons et perspectives"*, Ouagadougou : 27-35.
- DJITÉ, Paulin (2009) "Multilingualism: the Case for a New Research Focus", *International Journal of the Sociology of Language* 199: 1-7.
- EASTERLY, William (2001) "Can Institutions Resolve Ethnic Conflict?", *Economic Development and Cultural Change* 49: 687-706.
- EASTERLY, William and LEVINE, Ross (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics* 111: 1203-1250.

- FARDON, Richard and FURNISS, Graham (eds.). (1994). *African Languages, Development and the State*. New York: Routledge
- FEARON, James D. (2003) "Ethnic and Cultural Diversity by Country", *Journal of Economic Growth* 8: 195-222.
- FEARON, James D. and David LAITIN (2003) "Ethnicity, Insurgency, and Civil War", *American Political Science Review* 97: 75-90.
- FEDDERKE, Johannes; John LUIZ and DE KADT, Raphael (2008) "Using Fractionalization Indexes: Deriving Methodological Principles for Growth Studies from Time Series Evidence", *Social Indicator Resources* 85: 257-278.
- GADET, Françoise and VARRO, Gabrielle (2006) "Le «scandale» du bilinguisme", *Langage et Société* 2 (116): 9-28.
- GARDEAZABAL, Javier (2011) "Linguistic Polarization and Conflict in the Basque Country", *Public Choice* 149: 405-425.
- GINSBURG, Victor and WEBER, Shlomo (eds.). (2016). *The Palgrave Handbook of Economics and Language*. London: Palgrave Macmillan.
- GRÉVIN, Benoît (2016) "Le plurilinguisme, objet d'histoire", *Hypothèses* 19(1): 333-350.
- GUYOT, David (2007) "Plurilinguisme et métissage: le cas des métis du Togo", *International Journal of the Sociology of Language* 128: 73-94.
- HABYARIMANA, James; MACARTAN, Humphreys; POSNER, Daniel N. and WEINSTEIN, Jeremy M. (2007) "Why does ethnic diversity undermine public goods provisions?", *American Political Science Review* 101(4): 709-725.
- HADENIUS, Axel (1992). *Democracy and Development*. Cambridge: Cambridge University Press.
- HORNER, Kristine and WEBER, Jean-Jacques (2010) "Small Languages, Education and Citizenship: The Paradoxical Case of Luxembourgish", *International Journal of the Sociology of Language* 205: 179-192.
- JUILLARD, Caroline (2005) "Hétérogénéité des plurilinguismes en Afrique à partir du terrain sénégalais", *La Linguistique* 41 (2): 23-36.
- KLEIN, Carlo (2003) "La valorisation des compétences linguistiques sur le marché du travail luxembourgeois", CEPS/INSTEAD *Cahier PSELL* 139.
- LABART, Kelly (2010) "What is hidden behind the indicators of ethno-linguistic fragmentation?". Document de travail de la FERDI, Série "Indicateurs de développement innovants" / Idi10.
- LANE, Jan-Erik and ERSSON, Svante (1998) *Politics and Society in Western Europe*. 4<sup>th</sup> edition, London: SAGE.

- LA PORTA, Rafael; LOPEZ DE SILANES, Florencio; SHLEIFER, Andrei and VISHNY, Robert (1999) "The Quality of Government", *Journal of Law, Economics and Organization* 15: 222-279.
- LEESON, Peter T. (2005) "Endogenizing Fractionalization", *Journal of Institutional Economics* 1 (1): 75-98.
- LEWIS, M. Paul; SIMONS, Gary F. and FENNIG, Charles D. (eds.). (2013). *Ethnologue: Languages of the World*. Seventeenth edition, Dallas, Texas: SIL International. Online version: <<http://www.ethnologue.com>>.
- LUIZ, John M. (2015) "The impact of ethno-linguistic fractionalization on cultural measures: dynamics, endogeneity and modernization", *Journal of International Business Studies* 46: 1080-1098.
- MASELLA, Paulo (2011) "National Identity and Ethnic Diversity", *Journal of Population Economics* 26 (2): 437-454.
- MAURO, Paolo (1995) "Corruption and Growth", *The Quarterly Journal of Economics* 110 (3): 691-712.
- MONTALVO, José. G. and REYNAL-QUEYROL, Marta (2005) "Ethnic Polarization, Potential Conflict and Civil Wars", *American Economic Review* 95 (3): 796-816.
- NETTLE, Daniel (2000) "Linguistic Fragmentation and the Wealth of Nations: The Fischman-Pool Hypothesis Reexamined", *Economic Development and Cultural Change* 48 (2): 335-348.
- OCDE (2012). *Trouver ses marques: Les indicateurs de l'OCDE sur l'intégration des immigrés 2012*, Éditions. OCDE. <<http://dx.doi.org/10.1787/10.1787/9789264073432-fr>>.
- OKEDIJI, Tade O. (2005) "The dynamics of ethnic fragmentation: A proposal for an expanded measurement index", *American Journal of Economics and Sociology* 64(2): 637-62.
- POSNER, Daniel N. (2004) "Measuring Ethnic Fractionalization in Africa", *American Journal of Political Science* 48 (4): 849-863.
- RABUSHKA, Alvin and SHEPSLE, Kenneth A. (1972). *Politics in Plural Societies. A Theory of Democratic Instability*. Columbus, Ohio: Charles E. Merrill Publishing Company
- REILLY, Benjamin (2000-01) "Ethnic Fragmentation, and Internal Conflict: Confused Theories, Faulty Data, and the «Crucial Case» of Papua New Guinea", *International Security* 25 (3): 162-185.
- REILLY, Benjamin (2001). *Democracy in Divided Societies: Electoral Engineering for Conflict Management*. Cambridge: Cambridge University Press.

- REYNAL-QUEROL, Marta (2002) "Ethnicity, Political Systems, and Civil Wars2", *Journal of Conflict Resolution* 46(1): 29-54.
- REYNAL-QUEROL, Marta and MONTALVO, José G. (2000) "A theory of religious conflict and its effect on growth", *Instituto Valenciano de Investigaciones Económicas working paper* WP-EC 2000-04 (Deposito Lega: V-1710-2000)
- RODRIK, Dani (1998) "Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses", *NBER Working Paper Series Working Paper* 6350.
- ROOT, Maria P. P. (ed.). (1992). *Racially Mixed People in America USA*: SAGE Publications, Inc.
- ROOT, Maria P. P. (ed.). (1996). *The Multiracial Experience: Racial Borders as the Nex Frontier USA*: SAGE Publications, Inc.
- SCHOLAND, Andrew; TAUSCZIK, Yla R. and PENNEBAKER, James W. (2010) "Social Language Network Analysis", *Proceedings of the 2010 ACM conference on Computer supported cooperative work*: 23-26.
- STATEC. *Etat de la population, Population par sexe et par nationalité (x 1 000) 1981, 1991, 2001-2012*, <[http://www.statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=384&IF\\_Language=fra&MainTheme=2&FldrName=1](http://www.statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=384&IF_Language=fra&MainTheme=2&FldrName=1)>, consulté le 27/06/2012>.
- STAVEREN, Irene van and PERVAIZ, Zahid (2017) "Is it Ethnic Fractionalization or Social Exclusion, Which Affects Social Cohesion?", *Social Indicators Research* 130: 711–731.
- TAYLOR, Charles Lewis and HUDSON, Michael C. (1972). *World Handbook of Political and Social Indicators II*. second edition, New Haven: Yale University Press
- TEMPLE, Jonathan (1998) "Initial Conditions, Social Capital and Growth in Africa", *Journal of African Economies* 7 (3): 309-347.
- TRIM, John L. M. (2007) "Les langues vivantes au conseil de l'Europe 1954-1997. La coopération internationale en faveur de l'apprentissage tout au long de la vie, pour une communication efficace, un enrichissement culturel mutuel et la citoyenneté démocratique en Europe", *Division des Politiques linguistiques, Strasbourg*.
- TRUDELL, Barbara (2009) "Multilingualism in Sub-Saharan Africa: Describing a Phenomenon, Leveraging a Resource", article presented at the International Conference on Multilingualism, Bamako 19-21 January 2009.
- VIGDOR, Jacob L. (2004) "Community Composition and Collective Action: Analysing Initial Mail Response to the 2000 Census", *Review of Economics and Statistics* 86 (1): 303-312.

WANYONYI, Nakhisa Andrew (2012) "The East African Framework in the Context of Plurilingualism: A case of University Education in Regional Integration from St Augustine University of Tanzania focus", *Synergies Afrique des Grands Lacs* 2: 75-82.

WILLIAMS, Donald R. (2006) "The Economic Returns to Multiple Language Usage in Western Europe", IRISS Working Paper Series, 2006-07.