SUS.DIV – CEPR – PSE Conference

"Models of Cultural Dynamics and Diversity"

Paris School of Economics
11-13 December 2008

The Use of Diversity in Cultural Economics: Theoretical and Applied Considerations

*Françoise Benhamou (CEPN, University of Paris 13), Renato Flôres (Fundação Getulio Vargas), Stéphanie Peltier (Université de La Rochelle)

Conference Models of Cultural Dynamics and Diversity

Paris School of Economics, 11-13 December 2008

The Use of Diversity in Cultural Economics: Theoretical and Applied Considerations[⊕]

FIRST DRAFT

FRANÇOISE BENHAMOU*, RENATO G. FLÔRES JR** AND STÉPHANIE PELTIER***

* CEPN, University of Paris 13, France
**EPGE / Fundação Getulio Vargas, Brasil
*** GRANEM, University of Angers and University of La Rochelle, France

Abstract.

A growing literature especially investigates the determinants of long run cultural diversity, but this literature generally avoids the problem of defining and measuring diversity precisely. In this paper, we present a formal framework in order to measure diversity that takes into account the three dimensions of diversity (variety, balance, disparity). We discuss the relevance of two indexes (the Shannon-Wiener entropy index and the Stirling index), and we show that it is necessary to introduce a new index (HAI index) in order to better understand the scope of diversity in a given country. Then, we apply the different indexes to the measurement of diversity in the case of book edition (especially translations in literature) in 20 OECD countries from 1979 to 2005. We put forward some elements of interpretation of the differences in the relative level of diversity between those countries.

JEL Codes. L82, Z11
Key words. Cultural diversity; book industry; cultural economics

[⊕] We are indebted to Andrew Stirling for detailed comments on a previous version of section 2. Though they were partially taken into account, the authors remain solely responsible for the whole text to follow.

1. Introduction.

Cultural and social policies are more and more based on the very blurred concept of cultural diversity (Acheson and Maule, 2004). Analysts emphasize the role played by cultural diversity in economic development. A growing literature especially investigates the determinants of long run diversity. Even in the field of cultural activities, this literature generally avoids the problem of defining and measuring diversity precisely (see among others: Peterson and Burger, 1975; Burnett, 1992; Alexander, 1996; Lopes, 1992 and Dowd, 2004). Academic research makes the implicit hypothesis that diversity is desired *per se* (Dixit and Stiglitz, 1977).

Therefore, two questions arise: what should be called cultural diversity, and is it possible to build reliable indicators of diversity?

On the first point, let us admit that cultural diversity embraces as many categories as possible. Let us take the example of literature. Diversity designates the variety of author's origins, their genre, their religion, the diversity of their biographies, etc. Diversity may also designate the variety in the contents (length, style, etc.). In this paper, we emphasize an aspect of diversity: the linguistic origin of fiction books translated during a given period.

On the second point, and for each category of books (i.e. fiction, non-fiction, art, children's books, etc.), diversity has to be measured. First, if we make the hypothesis that consumers' utility grows with diversity, firms have an incentive to better understand and evaluate diversity in order to increase their own market share. Second, cultural diversity is a stake for book industry: if cultural policy aims to promote diversity, it is necessary to build indicators of the relative level of diversity generated by alternative strategies. An example of this question relies on the effects of subsidies or regulations in the book industry: do regulations (especially Retail Price Maintenance) contribute protecting book trade from market competition, or do they prevent diversity to be developed in the book industry (see especially Tirole, 1993 and Van Der Ploeg, 2004)? In spite of a wide range of categories and styles, and of a permanent increase in the number of new titles and editions, many papers denounce a propensity to standardization (Schiffrin, 2004).

In fact, variations in the definition and measure of diversity raise confusion in the cultural, economic and political debate. Empirical analysis needs strong grounds and cannot eliminate the question of the measure of diversity: the interpretation of the evolution and determinants of diversity depends on the way diversity is measured. We address this question from both a formal and conceptual viewpoints.

First, a formal appraisal of issues related to diversity indexes is made. Diversity cannot be only associated with the sheer multiplicity of types, forgetting that their relative frequencies are also crucial for defining this concept. This is the reason why we introduce the Shannon-Wiener entropy index as a reliable tool for measuring diversity. Moreover, following Stirling (1999), we show that it is necessary to take into account the degree of similarity between any given pair of objects or types, and we propose a second and complementary index. Moreover, we introduce an alternative index in order to introduce the distance between all the types or elements weighted by their proportion *and* the distance of each type or element with a referent.

In the second part,, we apply the different indexes to the measurement of diversity in the case of book edition (especially literature). We compare the evolution of the different indexes through time and between 20 OECD countries from 1979 to 2005. We make the hypothesis that a reliable (even non sufficient) criterion of diversity in literature is the number of books translated and the variety in the original languages of those books. The diversity of the origins of authors is a signal of the openness of a country to foreign cultures.

We also introduce the disparity between languages (Dyen *et al.*, 1992, Ginsburgh *et al.*, 2008). Following the work of Cavalli-Sfortza *et al.* (2000) and Ginsburgh *et al.* (2008), linguistic distance evaluations seem an objective way to approximate cultural distances[1], as Strorti (2001) and Sapiro (2008) demonstrate. Storti (2001) shows that languages reflect cultural differences. He explores how people from different cultures have different values, beliefs, ideas of good and evil, morality and immorality and shows that a wide part of those differences is reflected in the languages. He collects 51 cross-cultural dialogs and analyses, paying special attention to both the explicit and the implicit differences between Americans and people from Britain, France, and Germany. Among these groups, Americans and the French have the least in common, while Americans tend to have relative affinity with Germans[2]. In the same way, Sapiro (2008) emphasizes the role

[1] If measure of cultural distances exists (see Hofstede's surveys (1980, 1983), we choose to use the measure of linguistic distance for two major reasons: (a) cultural distances are available for countries, not for languages, while UNESCO data are for languages ; (b) cultural distances are more arbitrary. Based on statistical surveys of wide panels of IBM employees in more than 70 countries, cultural distances between nations are analysed through work-related values: power distance, individualism, uncertainty, avoidance, long-term orientation. Beyond the inherently low reliability of pool surveys, especially when they make international comparisons, questions and answers are biased in two ways. Firstly, the words do not mean the same thing in different countries (what appears to be the same question on individualism is unlikely to be truly equivalent in Japan and in the South of Europe). Secondly, poll surveys cannot grasp or take into account the entire complexity of cultural differences, in which history, religions, ways of life, and the social contract that is in force play a crucial role. In a nutshell, these surveys on cultural differences are more suitable for pointing out differences in job relations and corporate governance between different countries.

[2] e.g., Germany, like America, is more like a meritocracy, whereas France and Britain have an aristocratic history.

of the translations as a means in order to develop the dialog between cultures. The study of the origins of the foreign translated in the literature is a signal of the degree of cultural diversity supplied by publishers.

This concrete application illustrates the conceptual questions posed when we try to better understand and evaluate diversity. It raises new questions as well, dealing with the limits of indicators and the need for a strong theoretical and empirical framework. We emphasize the deep differences in the interpretation of the relative level and evolution of diversity according to the index selected for the same period and set of OECD countries.

2. Measuring diversity.

The present requirements on the (diversity) concept have made it mandatory to have a straightforward way of measuring the very diversity in concrete situations or markets. This led to the search of indicators or indexes to assess diversity. Supposing a suitably characterised context is given, basic elements for the construction of such indexes are a well-defined set of objects, outcomes or types, say 1, 2, …, n, and an associated frequency (or probability) distribution p_i , $1 \leq i \leq n$, $\sum_i p_i = 1$.

A common mistake, still present in many studies and arguments, is to associate diversity with the sheer multiplicity of types (variety), forgetting that their relative frequencies are also crucial for defining "the amount of diversity" (balance). In spite of different options duly taking into account the two basic constituents above, the Shannon-Wiener entropy index seems to be most favoured and, to many a number of viewpoints, the best candidate. Indeed, since Shannon (1948), several proofs of optimality of the entropy index have been produced. Its definition, as known, is:

$$H_{SW} = - \sum_i p_i \ln p_i \ , \qquad (1)$$

where, though in the theoretical developments the logs are assumed to be neperian, in practical applications they are often taken base 2.

Inspired perhaps in the works by Weitzman (1992, 1993), Stirling (1999) goes further and proposes a different index, taking into account the three dimensions of diversity: variety, balance and disparity. Disparity is defined as the degree of similarity between any given pair of objects or types. Stirling's proposal introduces a new element in the set of basic constituents, where objects, till then, were considered uniquely, and intrinsically, distinguished, no differences in their (relative) *proximities* being at stake. However, even

ecologists are aware of the 'redundancy hypothesis', related to different species with similar characteristics and, more importantly, similar functional roles. Systems with distinct H_{SW} would then show a nearly equivalent behaviour, as long as representatives of the same functional groups were present in both.

Cutting through more careful discussions on the problems raised by imposing a metric in objects' space, Stirling assumes the existence of a distance function d_{ij} , well-defined for all pairs (i,j). In this, we may also see an implicit influence of Lancaster (1966)'s early ideas – pioneered, in their turn, by Gorman (1953, 1956 and 1961) - to incorporate quality in consumer theory, where products – i.e., types – are defined by transformations of an original attribute's space[3]. In this way, a Euclidean distance can be naturally computed between products.

In the light of these assumptions, Stirling's proposal is:

$$H_{St} = \sum_{i,j} d_{ij}\, p_i\, p_j \quad . \qquad (2)$$

Distances between pairs of elements represent their mutual disparity (d_{ij}). Variety and balance can be captured by weighting distances by the product of the proportional importance in the system of each element in the pair ($p_i\, p_j$).

In its original formulation, the index is dependent on the measurement unit adopted for the distances, so that we prefer to impose a normalisation by setting the smallest distance, say d_{12} , equal to 1, and defining $d_{ij}^* = d_{ij} / d_{12}$, so that

$$H_{St}^* = \sum_{i,j} d_{ij}^*\, p_i\, p_j \quad , \quad d_{12}^* \equiv 1. \qquad (3)$$

Written as above, the index is invariant to linear transformations on the set of distances, though, annoyingly, it continues not to be invariant to other classes of transformations, even affine ones. Indeed, in spite of the fact that, in the cultural context, the more or less similarity among objects makes sense, use of (2), or (3), instead of (1) poses a few questions. While (1) enjoys important properties that aid in the interpretation of practical results, Stirling's idea presents a confusing behaviour. In his favour, though, it should be mentioned that his purpose in proposing (3) was exactly to point out how diversity can change with the different

[3] As known, purely economic approaches to diversity can differ. Rosen (2004), for instance, is an example of another independent line, though based on standard ideas on product differentiation and imperfect competition.

similarity relations among the units at stake. Indeed, in his view[4], Shannon's index bypasses this (important) question, by precisely assuming that the units occupy the vertices of a regular simplex in characteristics (or similarities) space.

With the above proviso in mind, two contrasting properties of the indexes to measure diversity will be discussed. The first point has to deal with the fact that Stirling Index does not depend of any referent. Stirling considers distances between each pair of elements: he does not introduce the distance between each element and one focal element that would play the role of such a referent. This approach may be relevant in order to measure diversity in some specific contexts, when the different types or elements are considered equivalent. By contrast, many economic problems deserve different approaches, and need the introduction of a referent. An example brings evidence of this need: we evaluate the level of diversity resulting of the introduction of new technologies by comparing this level with the one that prevailed in the context of the previous dominant technology. This is especially true in the case of literature. The diversity is not only measured by the proportion of translations among new publications, but also by the number of original languages, by the distance between those languages *and* by the distance between those different languages and the referent language (which is the one of the country in which we study the level of diversity).

Let us study a very mere example in order to illustrate this assertion. Let us suppose 4 types (a, b, c, e) with the same distribution of frequencies ($p_a = p_b = p_c = p_e = 0.25$). Distances between each pair of elements are : $d_{ab} = d_{bc} = d_{ae} = d_{ec} = 0.5$; $d_{ac} = 1$ et $d_{be} = 0.1$. The value of the Stirling Index is: $H_{st} = 0.194$. This value does not take into account any referent, i.e., in the case that we study *infra*, the distances between the translated languages and the language of the country analyzed and considered as a referent.

More precisely, let us imagine a country A where French is the dominant language. The structure of the literature book production would be: French: 20%, others (translated languages) 80% with: Czech: 20%, Slovak: 20%, English: 20%, German: 20%. Let us imagine a country B where German is the dominant language. The structure of the literature book production would be: German: 20%, others (translated languages) 80% with: Czech: 20%, Slovak: 20%, English: 20%, French: 20%. Both countries would have the same Stirling index, in spite of the proximity of the five languages in the case of Germany and of a wider distance between French and the others in the case of France. Stirling Index does not depend on the referent language. This is the reason why this index is not totally reliable. We need another index that reflects not only the proportions of original languages and their reciprocal distances, but also the openess of

[4] Personal communication by Andrew Stirling.

a specific country, with its own original language, its ability to translate distant languages, that reflect diverse original cultures.

This feature of the Stirling index is even more problematic when one studies, as we do in this paper, the diversity of translated languages. Let us imagine than both in France and in Germany the books translated by language are equally distributed among three languages: Czech, Slovak and English. Again, the Sirling Index will conclude to an identical level of diversity in both countries. In this case, the distances between translated languages and the referent language do not matter at all.

Therefore, we introduce an alternative index (H_{AI}) that takes into account the distance between all the types or elements (the languages in our case study) weighted by their importance (as Stirling does) *and* the distance of each type or element with the referent:

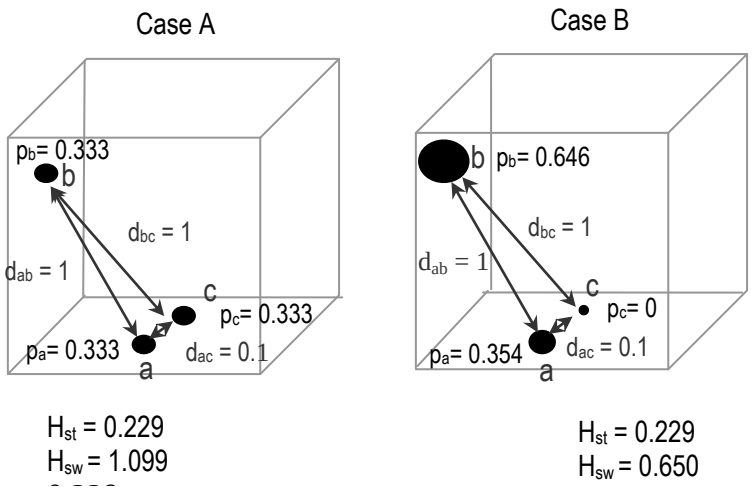
$$H_{AI} = \sum_{i,j} d_{ij}\, d_{ik}\, d_{jk}\, p_i\, p_j \quad \text{with } i,j \neq k \text{ and } k \text{ represents the referent}^5 \qquad (4)$$

The second point has to do with the range of indexes for a fixed number n of objects. The Shannon-Wiener entropy, as known, achieves its maximum for the uniform distribution, a result which has a strong intuitive appeal, as well as many practical and theoretical implications. Even for fixed distances, something equivalent cannot be stated for Stirling's or for our alternative index. As shown in the box 1, two situations, characterised by fixed distances between three types (a, b, c) but strongly different in the probability distribution, could lead to the same value of the $H_{st.}$ Although the Shannon-Wiener entropy without ambiguity shows that the diversity is higher in the case A, the use of Stirling's Index, as well as of our alternative one, does not lead to a clearcut conclusion.

In more detail, for values of number of types (n) greater than two, H_{St} 's maximum is difficult to interpret, given its dependence on the set of distances. It is easy to see that by varying the distances, a wide spectrum of maximum values can be attained by the index[6]. Moreover, for a given set of distances, the optimal probabilities may either be a corner solution, in the boundary of the simplex $\{ (p_1 , p_2 , p_3) \geq 0 \mid p_1 + p_2 + p_3 = 1 \}$, or indeed characterise a global maximum. Though in the latter case, the relative values of the probabilities follow some expected patterns, in both instances – particularly the former one - the solution bears a much less intuitive meaning.

[5] In our application, the dominant language of the translated country.
[6] For a proof of this statement in the cases n = 2 and n = 3, see Flôres (2006).

Box 1



Though nearly elementary, this may have disturbing consequences. On one hand, an important *and easy* reference value for assessing the diversity of a market or community with n types is lost. If we add the fact that the imposition of distances will be usually fraught with error, the meaning of the optimal frequencies for (2) (or (3) or (4)) becomes even shakier. On the other hand, if one changes the target and fixes the probability distribution, (3) (or (4)) may provide insights on the impact of the different metrics implied by the (different) classifications used.

3. Empirical Results and Discussion

In this part, we test the respective relevance of different indexes in the case of translations in literature. Proxies of diversity are not easy to define. We introduce a proxy that does not summarize all the dimensions of diversity, but that can be considered as a relevant signal of the openness of a given country to foreign cultures and authors. After a presentation of the data, we calculate the three indexes and show the contribution of the H_{AI} to a more reliable evaluation of diversity.

Data

The number of titles translated come from the UNESCO's database[7] that "contains cumulative bibliographical information on books[8] translated and published in about one hundred of the UNESCO Member States since 1979 and totals more than 1.800,000 entries in all disciplines: literature, social and human sciences, natural and exact sciences, art, history and so forth. UNESCO provides the general public with an irreplaceable tool for making bibliographical inventories of translations on a worldwide scale. International cooperation makes the "Index Translationum" a work tool that is unique in the world" [9]. Each year, data are sent by the bibliography centres or national libraries in the participating countries to the UNESCO Secretariat bibliographical data on translated books in all fields of knowledge.

However, as Ginsburgh *et al.* (2008) underline the UNESCO database is often criticized. Two main criticisms are usually addressed: (i) the definition of the book significantly differs among countries (some countries include doctoral dissertations, governmental, parliamentary and administrative documents, annual reports from firms, etc.)[10], and (ii) the data show sharp fluctuations. We partly avoid both criticisms, since (i) we focus only on the predominant gender "literature" for which there is more agreement on the definition, (iii) and on the twenty OECD countries for which the collect of data is the most reliable and updated and (iii) we also withdrawn some data that appear as outlyers. Note that Ginsburgh *et al.* (2008, p. 17), on a similar basis, choose to work on this data because, as Heilbron (1999) says, the "UNESCO source is the only international source that is readily available".

Our sample includes 20 countries over 27 years (1979-2005), which should have led to calculate 540 observations per index. However, for several countries (United-Kingdom and Sweden for instance) some data are missing. Finally, 419 observations have been calculated for each index. We select seventeen languages which represent the dominant language of the set of countries selected (see table 1). For each country, these languages always represent more than 80% of the literature books translated.

[7] Data were collected on UNESCO's website at the following address: http://databases.unesco.org/xtrans/stat/xTransXpert.a?lg=1.

[8] The scope of the data is limited to books. Periodicals, articles from periodicals, patents and brochures are not included.

[9] http://portal.unesco.org/culture/en/ev.php-URL_ID=7810&URL_DO=DO_TOPIC&URL_SECTION=201.html

[10] Moreover, this definition may change in a given country. For example, France has widened the scope of the definition in 2005, by considering that geographical maps, tourist guides, music partitions, etc. should be considered as books (with a specific VAT rate).

The measure of the distances between languages is based on the matrix of linguistic distances among Indo-European Languages proposed by Dyen *et al.* (1992)[11] (see table 1). Based on lexicographic methods, the distance between two languages *i* and *j* is equal to the percentage of words in the two languages which do not descend from a common world. This distance, normalised between languages, is comprised between 0 and 1. If the distance is close to 1, the two languages have completely different roots (for example, English and Japanese), and the distance is close to zero otherwise (like Slovak and Czech).

Table 1 - The Dyen Matrix of Linguistic Distances

	Ck	D	Dk	E	F	G	Gr	I	Ice	Po	Pol	Ru	S	Slo	Sw
Ck	0	0.762	0.746	0.759	0.773	0.741	0.836	0.753	0.766	0.764	0.234	0.255	0.760	0.126	0.767
D	0.762	0	0.337	0.392	0.756	0.162	0.812	0.74	0.408	0.747	0.769	0.776	0.742	0.769	0.308
Dk	0.746	0.337	0	0.407	0.759	0.293	0.817	0.737	0.221	0.750	0.749	0.740	0.750	0.732	0.126
E	0.759	0.392	0.407	0	0.764	0.422	0.838	0.753	0.454	0.760	0.761	0.758	0.760	0.750	0.411
F	0.773	0.756	0.759	0.764	0	0.756	0.843	0.197	0.772	0.291	0.781	0.778	0.291	0.765	0.756
G	0.741	0.162	0.293	0.422	0.756	0	0.812	0.735	0.409	0.753	0.754	0.755	0.747	0.742	0.305
Gr	0.836	0.812	0.817	0.838	0.843	0.812	0	0.822	0.802	0.833	0.837	0.832	0.833	0.832	0.816
I	0.753	0.740	0.737	0.753	0.197	0.735	0.822	0	0.755	0.227	0.764	0.761	0.212	0.749	0.741
Ice	0.766	0.408	0.221	0.454	0.772	0.409	0.802	0.755	0	0.763	0.758	0.754	0.763	0.757	0.211
Po	0.764	0.747	0.750	0.760	0.291	0.753	0.833	0.227	0.763	0	0.776	0.773	0.126	0.760	0.742
Pol	0.234	0.769	0.749	0.761	0.781	0.754	0.837	0.764	0.758	0.776	0	0.266	0.772	0.222	0.763
Ru	0.255	0.776	0.740	0.758	0.778	0.755	0.832	0.761	0.754	0.773	0.266	0	0.769	0.259	0.754
S	0.760	0.742	0.750	0.760	0.291	0.747	0.833	0.212	0.763	0.126	0.772	0.769	0	0.756	0.747
Slo	0.126	0.769	0.732	0.750	0.765	0.742	0.832	0.749	0.757	0.760	0.222	0.259	0.756	0	0.758
Sw	0.767	0.308	0.126	0.411	0.756	0.305	0.816	0.741	0.211	0.742	0.763	0.754	0.747	0.758	0

Notes. Since Finnish and Japanese are not Indo-European languages, they are not described in this table. Due to their linguistic remoteness, these languages have been considered as very far from each other and to every language in the table. The linguistic distance, in any cases, was set equal to one. Ck = Czech, D = Dutch, Dk = Danish, E = English, F = French, G = German, Gr = Greek, It = Italian, Ice = Icelandic, Po = Portuguese, Pol = Polish ; Ru = Russian , S= Spanish, Slo = Slovene, Sw = Swedish.

The contribution of the H_{AI} Index to diversity analysis

Does the H_{AI} Index lead to a similar or a different evaluation of diversity than the Shannon and the Stirling indexes? Table 2 provides the respective rank of the countries for each index, on the whole period. We observe (as expected) that the hierarchy between countries changes with the indexes, at least as far as the H_{AI} index is concerned. The Shannon Index ranking and the Stirling Index ranking are indeed nearly similar

[11] By comparison between 200 basic meanings, Dyen et al. estimated the linguistic distances for 95 Indo-European speech varieties (ie. languages and dialects). For our purpose, we only took the linguistic distances for the 15 dominant Indo-European languages of the 20 OECD countries of our sample. Since Finnish and Japanese are not Indo-European languages, they are not estimated. Given the strong difficulty to learn these languages for Indo-European population, we consider the linguistic distances between Japanese or Finnish to all other languages and between Japanese and Finnish to be maximum that is to say equal to one.

(the coefficient of correlation between these indexes for the whole sample is equal to 0.924). Hence, the mere introduction of the distances among translated languages does not change deeply the global rankings.

The study of the average ranking with the H_{AI} index leads to more significant changes. Table 2 shows that the average difference in the respective rankings is higher as soon as the H_{AI} index is taken into account. Hence, though positive and significant, the correlations between the H_{AI} index and the H_{SW} as well as the H_{ST} indexes are much lower (respectively 0.487 and 0.444). The most impressive differences in rankings concern Denmark, Finland, Iceland, Italy, Japan and Portugal. The case of the USA deserves an interpretation since with the H_{AI} index, the USA lose their leading position.

Table 2 – Average rankings and differences among rankings (1979-2005)

Countries	Average Ranking H_{SW} [1]	Average Ranking H_{ST} [2]	Average Ranking H_{AI} [3]	Differences of ranking between [1] and [2]	Difference of ranking between [1] and [3]	Difference of ranking between [2] and [3]
Austria	10.800	9.200	10.960	1.600	0.160	1.760
Czech Republic	8.000	8.308	5.462	0.308	2.538	2.846
Denmark	11.885	11.038	15.115	0.846	3.231	4.077
East Germany	3.455	2.091	4.636	1.364	1.182	2.545
Finland	7.577	11.154	1.654	3.577	5.923	9.500
France	11.346	11.462	8.500	0.115	2.846	2.962
Germany	9.308	8.885	10.885	0.423	1.577	2.000
Greece	4.333	3.571	1.429	0.762	2.905	2.143
Iceland	6.571	10.048	13.810	3.476	7.238	3.762
Italy	5.333	4.619	9.000	0.714	3.667	4.381
Japan	14.840	15.280	4.400	0.440	10.440	10.880
Mexico	14.333	11.556	11.667	2.778	2.667	0.111
Netherlands	11.440	11.120	13.480	0.320	2.040	2.360
Poland	7.462	7.654	6.346	0.192	1.115	1.308
Portugal	6.500	4.875	11.313	1.625	4.813	6.438
Slovak Republic	5.615	5.538	6.000	0.077	0.385	0.462
Spain	7.308	5.923	9.231	1.385	1.923	3.308
Sweden	13.182	13.909	14.818	0.727	1.636	0.909
United Kingdom	2.455	3.455	4.727	1.000	2.273	1.273
United States	1.042	1.333	2.667	0.292	1.625	1.333
Average difference between rankings				1.101	3.009	3.218
Correlation				0.923***	0.487***	0.444***

*** significant at the 1% level. The coefficients of correlation are calculated on the whole sample of indexes data and not on the average ranks (though results would have been very close).

How can we interpret these differences? We can stress the case of a few countries among those ones. In the case of Finland, English represents the most part of the translations, an average of 66.5% over the period. The other languages are obviously underrepresented, except Swedish and, to a lesser extent,

German and French. The rank obtained with the H_{AI} index results from the wide distance (and its evaluation) between the original language and the others (see table 1). Though Italy does not translate many languages of the sample (8 languages are non present among the translations, and 4 of them reach 1% only), the balance between the 4 more translated languages is somewhat higher (English 57%, German 10%, French 17%, Spanish 6%). But distances with Italian are less important, therefore, the H_{AI} index rank is lower than the rank of the two first indexes. Japan is rather weakly open to diversity (81.7% English, 7.7% French, 5% German, three languages at about 1% – Spanish, Italian and Russian – and all the rest minor than 1%): its rank is 15 with the two first indexes; nevertheless, the introduction of the distance with Japanese allows Japan to reach a much better rank for the H_{AI} index.

Hence, the H_{AI} Index introduces a more accurate evaluation of diversity. Diversity does not depend on the only distance between the different pairs of languages. Those three cases show the weight of the distance between the different languages *and* the domestic language of the country studied. This index helps understanding the openness of countries toward foreign *and* distant cultures and languages. The example of the USA bears evidence of this issue. The introduction of the distance with the American language and the other languages of the sample shows that the publishing sector is less open than what both the Shannon and Stirling approaches of diversity suggest. Indeed, the first rank of the USA according to both indexes lies in the balanced structure of translated languages: the dominant translated language – which is French – never exceeds 33% of market share.

The interpretation of H_{ST} and H_{AI} indexes is not straightforward

As highlighted in section 2, contrary to the Shannon-Wiener index of diversity, the interpretation of the H_{ST} and H_{AI} indexes are much less intuitive. Hence, two countries that obtain a close ranking according to the H_{AI} index may display very different characteristics.

Let us take the example of Finland and the USA. According to the H_{AI} index both countries are very close since their respective average rankings among the period are 1.65 and 2.67 (see table 2) and average indexes amount respectively to 0.150 and 0.123 (see table 4). However, when looking at the H_{SW} index a much more contrasted situation appears. The average ranking of the Finland is 7.58 vs. 1.04 for the US (average H_{SW} indexes for both countries are respectively 1.26 and 2.09). Indeed, in Finland, as stressed above, the translated literature is dominated by two languages: English (which a market share just below 70%) and Swedish. But those languages are considered as very distant from the Finnish. The case of the USA is strictly the opposite. The distribution of translated languages is much more balanced, but the

dominant languages are much closer from the English. Here is the explanation of the closeness of the H_{AI} index for both countries and their difference according to the H_{SW} index.

The evolution of cultural diversity with time

According to the Shannon-Wiener and the Stirling indexes, the level of diversity decreases significantly over the period, bearing evidence of a tendency toward more and more standardization and uniformity. The coefficient of correlation between both indexes and time is indeed significantly negative over the whole sample (see table 3). The fact that such an effect is not observed with the H_{AI} index could suggest that some countries have an increasing trend to translate some literature books which original language is more and more distant from their home language.

Table 3. Correlations between indexes and the number of titles translated, the population and time

Correlation coefficient	Hsw/titles	Hst/titles	H_{AI}/titles	Hsw/pop	Hst/pop	H_{AI} /pop	Hsw/time	Hst/time	H_{AI} /time
Total	-0.306***	-0.311***	-0.259***	0.409***	0.351***	0.308***	-0.307***	-0.295***	-0.070
Austria	-0.517***	-0.497***	-0.482**	0.600***	0.474**	0.454**	0.499***	0.354*	0.332*
Czech Rep.	0.011	-0.142	-0.406	-0.772***	-0.650**	-0.357	0.827***	0.715**	0.462
Danemark	-0.702***	-0.459**	-0.312	-0.943***	-0.516**	-0.306	-0.880***	-0.467**	-0.096
E. Germany	-0.546*	-0.163	0.853***				0.563	0.561	-0.551
Finland	-0.209	-0.296	-0.296	-0.760***	-0.738***	-0.737***	-0.587***	-0.559***	-0.559***
France	-0.170	0.215	0.502***	0.048	0.385*	0.551***	0.064	0.430**	0.621***
Germany	-0.299	-0.382	-0.410	-0.313	-0.294	-0.142	-0.346	-0.447	-0.354
Greece	-0.190	-0.417*	-0.423**	-0.308	-0.484**	-0.481**	-0.317	-0.525**	-0.527**
Iceland	0.425**	0.380*	0.338	-0.838***	-0.842***	-0.548***	-0.827***	-0.833***	-0.539***
Italy	-0.309	-0.388*	-0.265	-0.625***	-0.677***	-0.776***	-0.531**	-0.605***	-0.672***
Japon	-0.333*	-0.446**	-0.446**	-0.329*	-0.478**	-0.478**	-0.251	-0.383*	-0.383**
Mexico	-0.481	-0.532	-0.407	-0.263	-0.199	-0.407	0.006	-0.113	-0.036
Netherlands	-0.613***	-0.758***	-0.719***	-0.769***	-0.894***	-0.866***	-0.780***	-0.907***	-0.881***
Poland	-0.900***	-0.908***	-0.917***	-0.870***	-0.850***	-0.829***	-0.896***	-0.896***	-0.877***
Portugal	-0.091	-0.387	-0.640***	-0.015	-0.458*	-0.487**	0.123	-0.211	-0.677***
Slovak Rep.	0.000	0.288	0.626**	0.238	0.003	-0.038	0.382	0.030	-0.188
Spain	-0.193	-0.376*	-0.250	-0.635***	-0.682***	-0.488***	-0.663***	-0.730***	-0.528***
Sweden	0.567*	0.886***	0.896***	-0.498*	-0.784***	-0.703**	-0.298	-0.626*	-0.526*
UK	-0.222	0.226	0.011	0.607**	-0.402	0.417	0.712***	-0.185	0.611**
United States	0.413**	0.392**	0.431**	-0.732***	-0.793***	-0.731***	-0.712***	-0.781***	-0.711***

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level

Some countries display the same decreasing trend for the three indexes: Finland, Iceland, Netherlands, Poland, Spain and the US. In Poland, this decrease in diversity could be due to Transition that was

accompanied by a withdrawal in all the cultural activities, whatever their status (public or private) and their domain. In such a country, we observe a decline in all cultural consumptions, due to a rise in prices, a fall in subsidies, a new interest for the American cultural goods and services, etc. As for Spain, the explanation of the drop for the three indexes lies in the raise of English (from 55.7% in average market share over the period 1979-83 to 66.7% over the period 2001-2005) mainly at the expense of French. This concentration effect dominates the fact that English is more distant from Spanish than French. Let us now consider the US case. The decrease of the three indexes over time is due to an increase of French, the already dominant translation language (from 25.7% in average over the period 1979-83 to an average of 32.3% over the period 2001-2005), Spanish (from 9% to 13.7%) and to a lesser extent of Italian (from 6.5% to 9.9%), at the expense of more distant languages: Japanese (dropped from 7.9% to 1.8%) or Russian (from 11.4% to 8.8%).

Conversely, in France, for instance, diversity increases with time. The comparison of the different indexes for France is especially insightful. The fact that the H_{SW} index is independent from time whereas both other indexes increase with time suggests an increased market share for languages more different from the other translated languages and also more distant from French. Hence, whereas over the period 1979-1983, the average market share of English, German and Japanese were respectively 76.2%, 9.7% and 0.8%, these figures amount at 74.5%, 5.3% and 6.2% over the period 2001-2005. The raise of Japanese literature at the expense of the English and the German literature explains the increase of the H_{AI} index for France.

Discussion of the results

Though it is beyond the scope of this paper to study the determinants of the diversity indexes, we provide hereafter some elements of interpretation of the results. We analyze the potential links between diversity and the size of the country's population[12] on the one hand, and between diversity and the number of titles translated on the other hand.

In developed countries, like those belonging to our sample, the size of the population should affect positively the level of diversity. Indeed, the larger the size of the population, the larger niche markets constituted of readers interested in foreign literature very distant from the domestic one . Furthermore, minorities and immigrants – often over-represented in large Western countries – constitute another source of the development of profitable niche markets. All three diversity indexes are positively and significantly correlated

[12] The data on population comes from the World Bank database.

with the size of the population (see table 3), which suggests that, as expected, uniformization tends to be lower in large countries[13].

It turns out that the values of the diversity indexes are correlated to the number of titles translated from a foreign language (including languages that are not in our sample). Oddly enough, this correlation is negative and significant for the three indexes (see table 3).

First, let us consider the correlation on the whole sample. The negative correlation leads to qualify the high level of diversity of countries that translate a low number of titles of foreign literature. Indeed, as shown on table 4, for all indexes the four leading countries always translate a low number of titles. Hence, both the number of titles translated and the three diversity indexes contribute to the diversity analysis. Do notice that should the data on the whole number of titles of literature published in each country (domestic and translated) have been available, this correction of the diversity indexes by the number of titles translated would have been taken into account by the diversity indexes themselves.

Table 4 – Average diversity indexes, average number of titles translated and average population over the period.

Countries	Average H_{SW}	Average H_{ST}	Average H_{AI}	Nb. of titles translated	Population (in millions)
Austria	1.081	0.164	0.052	160	7,797
Czech Republic	1.145	0.154	0.081	1782	10,276
Denmark	0.953	0.147	0.029	1275	5,210
East Germany	1.984	0.276	0.111	480	
Finland	1.262	0.150	0.150	866	5,018
France	1.075	0.140	0.063	3349	56,908
Germany	1.165	0.160	0.050	4480	81,798
Greece	1.430	0.219	0.153	447	10,445
Iceland	1.317	0.155	0.037	198	0,258
Italy	1.313	0.198	0.057	861	57,086
Japan	0.766	0.103	0.103	1621	123,330
Mexico	0.941	0.146	0.046	159	80,996
Netherlands	1.084	0.150	0.037	1614	15,133
Poland	1.448	0.195	0.086	1036	37,830
Portugal	1.313	0.206	0.051	408	10,012
Slovak Republic	1.248	0.175	0.080	311	5,374
Spain	1.255	0.186	0.056	3690	39,234
Sweden	1.095	0.132	0.034	1116	8,424
United Kingdom	2.016	0.266	0.110	390	56,729
United States	2.094	0.274	0.123	490	255,472

[13] Do notice that this result still holds even if the USA are withdrawn from the sample.

Let us now consider the negative correlation between diversity indexes and the number of titles translated on a country basis (we focus on countries that obtain homogeneous results whatever the diversity index considered). We would expect that diversity increases when the number of titles translated grows. Yet, for countries like Austria, Netherlands or Poland, when translations are more numerous, some languages seem to grow in importance. We can assume, among different hypothesis, that it is due to learning effects that induce readers to be more interested in the foreign literatures that they have already discovered and learnt to understand. Familiarity would create a propensity to be interested in novels that are translated from close languages.

Conversely, the US case fits with expectations[14] (see table 3). Naturally, there is no effect of the domination of the English language among the languages translated in this country. The USA are not concerned by the polarization generally observed between American and domestic cultural products (Cohen and Verdier, 2008). Therefore, translations are spread among a relatively wide range of languages. Moreover, the peculiarity of the structure of the population, with a wide variety in the origins of the population, and an important proportion of the population whose native language is not English, is a strong ground for diversity.

4. Concluding remarks

Measuring cultural diversity is less easy than it may seem, demanding key conceptual decisions and careful statistical procedures. It is a stake for the implementation of cultural policies, but also in order to better understand the process of globalization. In this paper, we have shown that the combination of different indexes provides better information on the degree of openness of a country to different cultures. The H_{AI} Index, by taking into account variety, balance and distances not only among languages but also among languages and the domestic language seems to be a satisfying approach.

The paper emphasizes the only supply side. Further analysis should introduce the demand side: diversity in supply does not necessarily lead to the consumption of a wide diversity of goods and services. In spite of a struggle for 'higher diversity' in different contexts – in the movie and music markets, in the printing media, in the expression of local cultures/communities, in the right to post any kind of content in the web – there is no idea whatsoever on the desired levels of diversity. We still ignore the actual *level* desired for diversity. A disregard that must urgently be addressed. Complementary research is necessary in order to investigate the scope of demand for diversity. It should consist in the measure of the diversity consumed (Van der Wurff R. and J. Van Cuilenburg, 2001, Benhamou and Peltier, 2007) and address the correlation between the

[14] We ignore the case of the Sweden because of too many missing data for this country.

diversity supplied and the diversity consumed. More generally speaking, we can highlight the ignorance on optimal diversity levels. One may question the necessity to have an optimal, diversity value. We need the establishment of links between cultural diversity and measurable properties or services. Relevant further research would consist in the collect of data on the potential determinants of diversity. Beyond the influence of the size of the population and of time, we can underline: the level of economic development, the existence of ex-colonialist links, the share of a common frontier, the share of a common language, the geographical distance beween two countries, the proportion of immigrants, etc. Multiple regressions would help identifying the roots of diversity: one of the problems plaguing most applications is the lack of interpretative content of the indexes by themselves. In this vein, reference values – though not necessarily "optimal" ones - can be of great use.

5. References

Acheson K. and C. Maule, 2004, Convention on Cultural Diversity, *Journal of Cultural Economics*, 28(4): 243-256.

Alexander P.J., 1996, Entropy and Popular Culture: Product Diversity in the Popular Music Recording Industry, *American Sociological Review*, 61(1): 171-174.

Benhamou F. and S. Peltier, 2007, How Should Cultural Diversity be Measured? An Application using the French Publishing Industry, *Journal of Cultural Economics*, 31(2): 85-107.

Burnett R., 1992, The Implications of Ownership Changes on Concentration and Diversity in the Phonogram Industry, *Communication Research*, 19: 749-769.

Cavalli-Sfortza L., 2000, *Genes, Peoples and Languages*, North Point Press.

Cohen D. and T. Verdier, 2008, *La mondialisation immatérielle*, Conseil d'analyse économique, Rapport (76), Paris: La Documentation française.

Cowen T., 2002, *Creative Destruction,* Princeton: Princeton University Press.

Dixit A.K. and J.E. Stiglitz, 1977, Monopolistic Competition and Optimal Product Diversity, *American Economic Review*, 67(3): 297-308.

Dowd T.J., 2004, Concentration and Diversity Revisited: Production Logics and the U.S. Mainstream Recording Market, 1940 to 1990, *Social Forces*, 82: 1411-1455.

Dyen I., Kruskal J.B. and P. Black, 1992, An Indo-European Classification: a Lexicostatistical Experiment, *Transactions of the American Philosophical Society,* 82(5), Philadelphia: American Philosophical Society.

Flôres R.G., 2006, The diversity of diversity: further methodological considerations on the use of the concept in cultural economics. *Ensaios Econômicos* n° 626, Rio de Janeiro: EPGE/Fundação Getulio Vargas.

Ginsburgh V., Weber S. and S. Weyers, 2008, Economics of Literary Translation. A Simple Theory and Evidence, *Nota di Lavorro*, Fondazione Eni Enrico Mattei.

Gorman W. M., 1953, Community preference fields, *Econometrica,* 21: 63-80.

Gorman W. M., 1956, A possible procedure for analysing quality differentials in the egg market, Mimeo, Ames, Iowa State College.

Gorman W. M., 1961, On a class of preference fields, *Metroeconomica,* 13: 53-6.

Heilbron J., 1999, Towards a sociology of translation, Books translations as a cultural word system, *European Journal of Social Theory,* 2: 429-444.

Hofstede G., 1980, *Culture's Consequences: International differences in work-related values*, California: Sage Publications.

Hofstede G., 1983, National cultures in four dimensions: a research-based theory of cultural differences among nations, *International Studies of Management and Organization*, 13 (1-2): 46-74.
Lancaster K. J., 1966, A new approach to consumer theory, *Journal of Political Economy,* 74: 132-57.
Lopes P.D., 1992, Innovation and Diversity in the Popular Music Industry, 1969 to 1990, *American Sociological Review*, 57(1): 56-71.
Peterson R.A. and D.G. Berger, 1975, Cycles in Symbol Production: the Case of Popular Music, *American Sociological Review,* 40(2): 158-173.
Peterson R.A. and D.G. Berger, 1996, Measuring Industry Concentration, Diversity, and Innovation in Popular Music, *American Sociological Review*, 61(1): 175-178.
Rosen S., 2004, *Markets and Diversity*, Cambridge: Harvard University Press.
Sapiro G., 2008, *Translatio. Le marché de la traduction en France à l'heure de la mondialisation*, Paris: CNRS Editions.
Schiffrin A., 2004, *Le contrôle de la parole – L'édition sans éditeurs (suite)*, Paris: La Fabrique.
Shannon C.E., 1948, A mathematical theory of communication, *Bell Systems Technical Journal,* 27: 379-423, 623-56.
Stirling A., 1999,On the economics and analysis of diversity, Mimeo, *SPRU Electronic Working Paper,* n. 28.
Storti C., 2001, *Old World, New World: Bridging Cultural Differences: Britain, France, Germany, and the U.S*, Intercultural Press.
Tirole J., 1993, *Théorie de l'organisation industrielle*, Paris: Economica.
Van der Ploeg F., 2004, Beyond the Dogma of the Fixed Book Price Agreement, *Journal of Cultural Economics*, 1: 1-20.
Van der Wurff R. and J. Van Cuilenburg, 2001, Impact of Moderate and Ruinous Competition on Diversity: the Dutch Television Market, *Journal of Media Economics*, 14 : 213-229.
Weitzman M.L., 1992, On diversity, *Quarterly Journal of Economics,* 107(2): 363-406.
Weitzman, M.L., 1993, What to preserve ? An application of diversity to crane conservation, *Quarterly Journal of Economics,* 108(1): 157-83.

Appendix

Shannon Index (H_{sw})

Countries	Average	Standard deviation	Median	min	max
Austria	1,081	0,381	1,046	0,721	1,724
Czech Republic	1,145	0,159	1,107	1,016	1,349
Denmark	0,953	0,534	1,188	0,456	1,337
East Germany	1,984	0,032	1,980	1,923	2,056
Finland	1,262	0,026	1,295	0,928	1,482
France	1,075	0,054	1,078	0,880	1,419
Germany	1,165	0,028	1,156	0,958	1,581
Greece	1,430	0,384	1,424	0,972	1,785
Iceland	1,317	0,352	1,332	0,905	1,568
Italy	1,313	0,020	1,317	1,132	1,475
Japan	0,766	0,173	0,747	0,568	1,039
Mexico	0,941	0,073	0,928	0,641	1,175
Netherlands	1,084	0,126	1,070	0,810	1,688
Poland	1,448	0,658	1,104	0,963	2,147
Portugal	1,313	0,026	1,305	1,196	1,425
Slovak Republic	1,248	0,174	1,280	1,093	1,474
Spain	1,255	0,053	1,254	1,102	1,429
Sweden	1,095	0,094	1,089	0,950	1,201
United Kingdom	2,016	0,106	2,018	1,927	2,099
United States	2,094	0,224	2,104	1,836	2,273

Stirling Index (H_{st})

Countries	Average	Standard deviation	Median	min	max
Austria	0,164	0,038	0,155	0,113	0,249
Czech Republic	0,154	0,015	0,147	0,140	0,178
Denmark	0,147	0,003	0,145	0,118	0,173
East Germany	0,276	0,002	0,275	0,270	0,282
Finland	0,150	0,003	0,153	0,106	0,177
France	0,140	0,010	0,143	0,106	0,183
Germany	0,160	0,007	0,160	0,128	0,225
Greece	0,219	0,051	0,222	0,184	0,256
Iceland	0,155	0,061	0,153	0,102	0,189
Italy	0,198	0,008	0,199	0,166	0,223
Japan	0,103	0,038	0,099	0,075	0,157
Mexico	0,146	0,005	0,147	0,089	0,188
Netherlands	0,150	0,059	0,144	0,105	0,219
Poland	0,195	0,096	0,146	0,120	0,285
Portugal	0,206	0,022	0,211	0,181	0,226
Slovak Republic	0,175	0,012	0,174	0,142	0,203
Spain	0,186	0,008	0,187	0,161	0,212
Sweden	0,132	0,075	0,142	0,035	0,152
United Kingdom	0,266	0,002	0,265	0,258	0,278
United States	0,274	0,042	0,277	0,230	0,295

Alternative Index (H_{AI})

Countries	Average	Standard deviation	Median	min	max
Austria	0,052	0,011	0,049	0,036	0,079
Czech Republic	0,081	0,003	0,079	0,075	0,090
Denmark	0,029	0,007	0,028	0,021	0,043
East Germany	0,111	0,008	0,112	0,100	0,119
Finland	0,150	0,003	0,153	0,106	0,177
France	0,063	0,013	0,061	0,047	0,091
Germany	0,050	0,002	0,049	0,039	0,075
Greece	0,153	0,035	0,156	0,129	0,179
Iceland	0,037	0,014	0,036	0,024	0,051
Italy	0,057	0,004	0,056	0,046	0,066
Japan	0,103	0,038	0,099	0,075	0,157
Mexico	0,046	0,004	0,046	0,031	0,061
Netherlands	0,037	0,014	0,034	0,027	0,056
Poland	0,086	0,022	0,079	0,063	0,111
Portugal	0,051	0,012	0,050	0,041	0,066
Slovak Republic	0,080	0,000	0,079	0,066	0,091
Spain	0,056	0,002	0,056	0,047	0,065
Sweden	0,034	0,023	0,037	0,004	0,042
United Kingdom	0,110	0,006	0,111	0,097	0,118
United States	0,123	0,016	0,125	0,102	0,142