Working papers series

WP ECON 17.07

Ranking Languages in the European Union: Before and After Brexit

Victor Ginsburgh
ECARES, Universite Libre de Bruxelles and CORE

Juan D. Moreno-Ternero
Department of Economics, Universidad Pablo de Olavide & CORE

Shlomo Weber
Department of Economics, Southern Methodist University, New Economic School, and Department of Economics, Lomonosov Moscow State University

Keywords: Ranking methods, European Union, Communicative benefits, Linguistic disenfranchisement, Official languages

JEL Classification: C78, D61, D63

Department of Economics

Ranking Languages in the European Union: Before and After Brexit[1]

Victor Ginsburgh
ECARES, Université Libre de Bruxelles, Belgium,
and CORE, Université catholique de Louvain, Belgium

Juan D. Moreno-Ternero
Department of Economics, Universidad Pablo de Olavide, Seville, Spain,
and CORE, Université catholique de Louvain, Belgium

Shlomo Weber
Department of Economics, Southern Methodist University, USA,
New Economic School, Moscow, Russia,
and Department of Economics, Lomonosov Moscow State University, Russia

November 29, 2016

Abstract

This article presents a framework for evaluation of the impact of languages in multilingual societies. We consider several ranking methods based on various principles, including *minimal disenfranchisement*, *communicative benefits*, *utilitarianism*, and the game-theoretical concept of the *Shapley Value*. We use data from a Special Barometer survey to apply these methods to languages within the European Union and conclude that they generate quite consistent results. Finally, we analyse the impact of *Brexit* on the rankings, especially in the case where English forfeits its status as an *official* language of the Union.

[1]The first author is grateful to Michel Vanden Abeele for clarifications on the status of English if Brexit takes place and to Israel Zang for unending discussions on Shapley ranking. The second author acknowledges financial support from the Spanish Ministry of Economics and Competitiveness through the research project ECO2014-57413-P. The third author wishes to thank the Russian Science Foundation for its financial support through the research project $\#15-18-00098$. The detailed comments by seminar participants at UMASS, Amherst, two anonymous referees, as well as an associate editor of this journal are gratefully acknowledged.

1 Introduction

Multilingualism is a pervasive phenomenon that is traced back to ancient times. Members of small language communities were obliged to speak two or more languages for purposes of trade and other business or social interaction outside their own town or village (Holden, 2016). In more recent times, globalization and cultural openness foster multilingualism. A point in case is Europe, where there is no predominant language and English is often used as a language of communication. However, in multilingual countries such as Belgium (French, Dutch and German), Switzerland (German, French, Italian and Romanche), Spain (Spanish, Catalan, Basque and Galician) or Finland (Swedish and Finnish), it is usual to master two or even more languages. Some languages such as Danish, Swedish and Norwegian are close enough that it is generally more common for people to use their mother tongue rather than English in meetings.

Over the last twenty years it has been shown that institutional wastefulness, bureaucratic inefficiency, corruption, political instability and slow economic growth may result from excess of linguistic or ethnic diversity (see surveys in Alesina and La Ferrara, 2005, and Ginsburgh and Weber, 2011). In response, many multilingual countries found necessary to reduce the number of languages that were spoken and select a unique or a small number of languages to be used for various purposes by all citizens. This process started a long time ago in several European countries (such as France, the United Kingdom, Russia and Spain), and more recently in South Asia (e.g., India, Sri Lanka, and others).[2] A transition from the status quo with many languages to a situation with a smaller number of official languages could be reasonably smooth (e.g., the United Kingdom). In other cases, a feeling of disenfranchisement by large parts of the populations who did not speak any of the chosen languages, led to violence and even long civil wars (e.g., Sri Lanka).

English was the official language of the area that now comprises India and Pakistan under British rule, which ended in 1947. Nowadays, some 450 languages are alive in India and some 300 dialects and languages are spoken in Pakistan. The Pakistani constitution, which was passed in 1973, included a clause specifying that the government must make Urdu (far less spoken within the country than Punjabi) the national language within 15 years. The clause had not been enforced but, in 2015, the government of Pakistan announced plans to make Urdu the sole official language and abolish English as the second official language. Despite a similar language clause in its constitution, India continues to

[2]See, for example, Wright (2016) for Western Europe, Kadoshnikov (2016) for Russia and Sonntag (2016) for South Asia.

use both English and Hindi as its official languages. Actually, some fifty years ago, India attempted a three-language formula that required each citizen to speak English, Hindi and another local language. A similar approach was introduced in Nigeria, the most populous African country with 141 million inhabitants who speak over 500 languages and are divided into 250 ethnic groups. The three-language formula, based on the use of Hausa, Yoruba, and Igbo, was considered a unifying device for this diverse country. However, in both cases this implementation failed, due to the lack of qualified instruction and the resistance of linguistic groups to learn another language.

While linguistic frictions still exist in Belgium, Spain, and some Eastern European countries, one can safely argue that transition was, to a large extent, successful in Europe. This might be one of the reasons why multilingualism is one of the foundations of the European Union (EU). In the EU, every member country has the right to recognize one of its official languages as an official language of the EU. Regulation 1/1958 of the Treaty of Rome, signed in 1958, was extended to all new members who joined the EU through the successive enlargements. Today, the EU consists of 28 countries and 24 languages.[3] This results in a rather large translations budget (over one billion Euros/year), as all legal documents have to be translated in all the official languages, and Members of the European Parliament have the right to use their mother tongue, which needs a large host of interpreters.[4] Still, no citizen of any country (with the exception of children at school) is forced to learn a non-native language spoken in the EU.

We can infer from the above discussion that, in many situations, it is necessary to select a language, or a group of languages, from a larger set. Though the reason to reduce the number of languages is usually grounded on economic considerations (fostering integration, increasing interregional trade, saving on the cost of translation or interpretation, easing communication and education), it is often based on subjective and/or political reasons. In this paper, we present a stylized framework to rank languages in multilingual societies. The techniques we describe are based on normative grounds and result in five computable rankings, that are useful for decision-making in choosing language(s) to be considered as working, educational, official or otherwise approved set of languages.

[3]Some languages are official in more than one country: Dutch in Belgium and The Netherlands, English in the United Kingdom, Ireland and Malta, German in Germany, Austria and Belgium, and French in France and Belgium.

[4]Fidrmuc, Ginsburgh, and Weber (2007), Ginsburgh, Ortuno-Ortin and Weber (2011) and Ginsburgh and Weber (2005, 2011) suggested scenarios in which the number of official languages could be curtailed by using methods with some reminiscence to the ones discussed in the paper.

The first principle we examine, that of *Minimal Disenfranchisement*, focuses on the number of individuals who do not speak any of the chosen languages.[5] It is equivalent to Van Parijs' (2011) *Minimex* principle of minimal exclusion, which searches for a language that guarantees that the number of individuals who do not speak this language is the smallest possible. The notion of linguistic disenfranchisement, which may arise when the linguistic rights of parts of the society are denied or restricted, may have far-reaching consequences for economic growth and political stability and should be treated very carefully.

Another principle is that of *Communicative Benefits*, introduced by Selten and Pool (1991).[6] Note that, while the principle of Minimal Disenfranchisement does not distinguish between perfect and poor command of the language, as long as an individual speaks it, the communicative benefits approach is based on the perfect command of a language. It asserts that the utility of an individual is positively correlated with the number of others with whom she can perfectly communicate. Indeed, trade opportunities, earnings and job prospects could be enhanced by the ability of individuals to communicate with each other. The opportunity to learn about different cultures becomes more important in the globalized world we live in. In our context, it implies that a language is "superior" to another if it allows more pairs of individuals to communicate perfectly. It is easy to see that the highest level of communicative benefits obtains by choosing the language with the largest number of individuals who know it perfectly.

The two principles focus on the extreme cases of language knowledge, where individuals either do not know it at all or speak it perfectly. The *Aggregate Knowledge* principle would allow to extend this dichotomy to intermediate cases of language knowledge and postulates that a language is preferred to another if its aggregate knowledge (weighing by individual degrees) is higher than that of another language. This is essentially the principle underlying *Utilitarianism*, a deeply rooted notion in economics and philosophy, which can be traced back to Jeremy Bentham and John Stuart Mill.

These three principles would trivially lead to the same ranking under the premise of dichotomous knowledge of languages, in which agents either speak or do not speak a certain language. In a more general setting, which allows for intermediate knowledge levels, we show that the principles may yield different rankings.

We also consider two game-theory-based rankings by using the *Shapley Value*, the

[5]See Ginsburgh and Weber (2005) and Ginsburgh et al. (2005) for the discussion on linguistic disenfranchisement, with a special emphasis on the EU.

[6]See also Ginsburgh et al. (2007), Ginsburgh and Weber (2011), and Athanasiou et al. (2016).

well-known solution concept for cooperative games. In our setting, this concept amounts to the following. Suppose each agent speaks a set of languages and assigns, in some predetermined way, weights to each of them. These weights could be either equal, or proportional to the levels of knowledge. Adding these weights across agents produces a score for each language and, thus, their ranking. As we show later, the two ensuing rankings based on the Shapley value can lead to rankings that are fundamentally different from the previous three.

The five methods with their normative foundations are described in more detail in Section 2. In Section 3, we use data from a Special Eurobarometer survey and apply these methods to rank languages spoken in the (pre-Brexit) EU. The results, somewhat surprisingly, indicate that, in spite of their different normative grounds, all rankings yield similar results. This indicates that the ranking of languages within the EU is more robust than what might be thought. Section 4 uses *Minimal Disenfranchisement*, *Communicative Benefits* and *Aggregate Knowledge* to study the consequences of Brexit. Section 5 concludes.

2 Defining and analyzing rankings

We describe the linguistic landscape of a given society by a *language matrix*, A, whose rows refer to citizens (*agents*), and its columns to *languages*. Each entry a_{ij} of the matrix denotes the knowledge level of language j by agent i. We make the normalizing assumption that $a_{ij} = 0$ reflects no knowledge whatsoever, $a_{ij} = 1$ reflects perfect knowledge, and also assume that there exists a finite set of intermediate levels of knowledge ranging from 0 to 1. Here is an example of such a matrix.

Example 1

$$A = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 1/2 \\ 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 1 & 1/2 \\ 1 & 1/2 & 1/2 \end{pmatrix}.$$

This matrix refers to the case in which there are eight agents and three languages. Agents 1, 2 and 3 only have intermediate knowledge of languages 2 and 3, which we model by the number 1/2, and no knowledge of language 1. Agents 4 and 5 have perfect knowledge of 1, intermediate knowledge of 3 and no knowledge of 2. Agent 6 only has perfect knowledge of 2, and does not know any other. Agent 7 has perfect knowledge of 2, intermediate knowledge of 3 and no knowledge of 1. Finally, agent 8 has perfect knowledge of 1, and intermediate knowledge of the other two languages.

2.1 Definitions

We now introduce our ranking methods. Formally, let N denote the set of agents, and L the set of spoken languages in the society.

Minimal Disenfranchisement (MD). Speakers of various languages are accounted for, whether they speak the language perfectly or not, as long as they speak it to some degree. Proficiency is thus disregarded, and all positive entries in matrix A are replaced by ones. We define the *Minimal Disenfranchisement* (MD) ranking as the one that ranks languages according to the number of disenfranchised individuals, i.e., those who do not know them.[7] The larger the number of disenfranchised agents, the lower the ranking of the language. Thus, according to MD, language j is ranked higher than language k if $\#\{i \in N | a_{ij} = 0\} < \#\{i \in N | a_{ik} = 0\}$. In Example 1, MD thus ranks language 3 first (7 speakers), then language 2 (6 speakers) and finally, language 1 (3 speakers).

Communicative Benefits (CB). The entries a_{ij} that represent the degree of knowledge of a language can be bounded from below by imposing $a_{ij} \geq \underline{a}_i$ and $a_{ik} \geq \underline{a}_k$, where both $\underline{a}_i$ and $\underline{a}_k$ take values between 0 and 1. If both bounds are set to 1, only perfect or native speakers are counted. This defines the *Communicative Benefits* (CB) ranking. Thus, according to CB, language j is ranked higher than language k if $\#\{i \in N | a_{ij} = 1\} > \#\{i \in N | a_{ik} = 1\}$. This is equivalent to ranking languages according to the number of pairs that could communicate perfectly in each language. In Example 1, CB ranks language 1 first (3 perfect speakers), language 2 second (2 perfect speakers), and finally, language 3 last (0 perfect speakers).

[7] We may extend the notion of disenfranchisement to include those who speak languages only superficially.

Aggregate Knowledge (AK). If both $\underline{a}_i$ and $\underline{a}_k$ are set to $\varepsilon > 0$ sufficiently small, all agents who know the language, even if their knowledge is not perfect, are accounted for. But we count them weighing by their knowledge levels. This defines the *Aggregate Knowledge* (AK) ranking. In Example 1, language 1 scores 3, language 2 scores 4, and language 3 scores 3.5. Language 2 is thus first, language 3 is second, and language 1 is third. More generally, language j is ranked higher than k if $\sum_{i \in N} a_{ij} > \sum_{i \in N} a_{ik}$. AK ranks languages according to their aggregate knowledge level across society.

The first two rankings can be interpreted as translations to our context of *Approval Voting* (see Brams and Fishburn, 1978). This method allows each voter to cast her vote for as many candidates she wishes; each positive vote is counted in favour of the candidate. The votes are then added by candidate, and the winner is the one who gets the largest number of votes. All other candidates can also be ranked, according to the number of votes they obtain. *Minimal Disenfranchisement* obtains when an agent *approves* all languages with partial knowledge ($a_{ij} > 0$), whereas the *Communicative Benefits* obtains when an agent *approves* all languages with perfect knowledge only ($a_{ij} = 1$).

The third ranking can be thought of generalizing *Approval Voting*, to *Range Voting* (see Smith, 2004) or *Evaluative Voting* (see Hillinger, 2005).[8] Here, agents can express further evaluations of alternatives (for example numbers between 0 and 1) beyond just approving (or disapproving) them.

An alternative to Approval Voting is *Cumulative Voting* (Glasser, 1959; Sawyer and MacRae, 1962). It allows voters to distribute points among candidates in any arbitrary way. An interesting case is the one in which every agent is endowed with a fixed number of votes that are evenly divided among all candidates for whom she votes. This would translate into our context as the *Shapley ranking*.[9] Formally:

Shapley (S). Let $L_i(A) = \{j \in L : a_{ij} > 0\}$ denote the set of languages agent $i \in N$ has some knowledge of, and let $l_i(A)$ denote the number of such languages. Then, we say that

[8]The difference between the methods actually lies on whether alternatives (knowledge levels in our setting) can take any value in $[0, 1]$ or only a finite number of values.

[9]The name of the ranking comes from the rule introduced by Ginsburgh and Zang (2003) for the so-called museum pass game. Bergantiños and Moreno-Ternero (2015) extend the rule to more general museum pass problems. See also Ginsburgh and Zang (2013), who apply the rule for the ranking of wines.

language j is ranked higher than k if

$$\sum_{i\in N, j\in L_i(A)} \frac{1}{l_i(A)} > \sum_{i\in N, k\in L_i(A)} \frac{1}{l_i(A)}.$$

Consider the following matrix in which only the two extreme levels of knowledge are present.

Example 2

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

Matrix A describes a situation with three agents and four languages. Agent 1 only knows language 1, while the two other agents know languages 2, 3 and 4. Then, as stated above, the first three ranking methods discussed earlier would endorse the ranking in which languages 2, 3 and 4 come first, and language 1 comes last. The Shapley ranking inverts the ranking. Indeed, language 1 obtains a score of 1, while each of the three others obtains 2/3 (1/3 for each voter times 2 as there are two voters) and are thus ranked after language 1.

In order to interpret Shapley rankings, one could imagine that each agent is endowed with one vote which is shared among the candidates for whom she votes. Shares for each candidate are then added as above, and the candidate who gets the largest number of shares wins. All other candidates are ranked accordingly. As shown by Ginsburgh and Zang (2003), these numbers represent *Shapley Values* of candidates, which can also be interpreted as their *powers* resulting from the vote. This is different from Approval Voting where voters can cast as many votes as they wish.

To conclude with the inventory of rankings, one could consider a reasonable generalization of the Shapley ranking, in which agents distribute points among languages proportionally to their knowledge levels, which we call *Weighted Shapley* ranking. Formally,

Weighted Shapley (WS). Language j is ranked higher than k if

$$\sum_{i\in N, j\in L} \frac{a_{ij}}{\sum_{h\in L} a_{ih}} > \sum_{i\in N, k\in L} \frac{a_{ik}}{\sum_{h\in L} a_{ih}}.$$

Obviously, (S) and (WS) would coincide for cases in which knowledge is only maximal or minimal (such as in Example 2). Consider now the following matrix with four agents and three languages:

Example 3

$$A = \begin{pmatrix} 1 & 1/4 & 3/4 \\ 0 & 1/4 & 1 \\ 0 & 1 & 0 \\ 1/4 & 0 & 0 \end{pmatrix}.$$

The Shapley (S) scorings are as follows: Language 1 gets 1/3 from agent 1 and 1 from agent 4, This totals 4/3. Language 2 gets 1/3 from agent 1, 1/2 from agent 2, and 1 from agent 3 which totals 11/6. Language 3 gets 1/3 from agent 1 and 1/2 from agent 2 which totals 5/6. Therefore, language 2 would come first, followed by language 1 and then language 3.

The Weighted Shapley (WS) scorings would come as follows: Language 1 gets 1/2 from agent 1 and 1 from agent 4 which totals 3/2. Language 2 gets 1/8 from agent 1, 1/5 from agent 2, and 1 from agent 3 which totals 53/40. Language 3 gets 3/8 from agent 1 and 4/5 from agent 2 which totals 47/40. This produces the following ranking: language 1 would come first, followed by 2 and then 3.

2.2 A normative analysis

We now consider a number of reasonable assumptions (axioms). As we shall see, different combinations of these axioms will characterize each of the five rankings presented above.

Axiom 1 High Invariance. The ranking is not altered if positive non-perfect knowledge levels are replaced by perfect knowledge level, that is if every $0 < a_{ij} < 1$ in matrix A is replaced by 1.

Axiom 2 Low Invariance. The ranking is not altered if non-perfect knowledge levels are replaced by no-knowledge levels, that is if every $a_{ij} < 1$ in matrix A is replaced by 0.

Axiom 3 Impartiality. If all *but two* agents in a society have the same knowledge of two languages j and j', with the exception of agents i and i', and if i knows language j perfectly and has no knowledge of j', while the reverse holds for i', then both languages j and j' are ranked at the same level.

Axiom 3' Generalized Impartiality. Let the set of knowledge levels of every language and every agent be finite. If agent i increases her knowledge of language j to an immediate higher level, while agent i' decreases her knowledge of j to the immediate lower level, the

ranking of languages does not change.[10]

Axiom 4 (Strong) Pareto Optimality. If all agents speak language j at least as well as language j', with at least one agent speaking j strictly better than j', then j is strictly ranked above j'.

We obtain the following characterizations for rankings MD, CB and AK.[11]

Proposition 1 *MD is the unique ranking that satisfies Axioms 1, 3 and 4.*
Proposition 2 *CB is the unique ranking that satisfies Axioms 2, 3 and 4.*
Proposition 3 *AK is the unique ranking that satisfies Axioms 3' and 4.*

We chose to focus on our specific case to provide new characterization results for our first three rankings. Note that some of the many existing characterizations of approval voting (see, for instance, Xu, 2010) can be used for the *Minimal Disenfranchisement* and *Communicative Benefits* rankings. Likewise, some of the many existing characterizations of *utilitarian* criteria (see, for instance, Blackorby et al. 2002) can be used for the *Aggregate Knowledge* ranking.

For the *Shapley* rankings, we consider a different scenario, previously used by Bergantiños and Moreno-Ternero (2015).[12] More precisely, suppose that each agent has a *vote* to be allocated among the languages she speaks. The ranking would then be generated from the overall number of votes allocated to each language. We consider the next three axioms for such a context.

Axiom 5 Equal Treatment of Equals. If two languages have the same number of speakers, then they should receive the same number of votes.
Axiom 6 Dummy. If no agent speaks a language, then it should get no votes.
Axiom 7 Additivity. Given two groups of speakers, it is equivalent to consider them separately, or as the same group.

This leads to Proposition 4, which can be stated as follows:[13]

[10] This is equivalent to the axiom recently introduced by Macé (2015), under the term *Compensation*, in a setting of "voting with evaluations".

[11] Appendix 1 contains more formal statements of the axioms, as well as the proofs of the propositions.

[12] We are not aware of any axiomatic characterization of (Equal and Even) Cumulative Voting. Both Approval Voting and Cumulative Voting can, however, be seen as members of a family of voting procedures dubbed as *Size Approval Voting*, which are characterized by Alcalde-Unzu and Vorsatz (2009).

[13] See Appendix 1 for the formal proof.

Proposition 4 *S is the unique ranking that satisfies Axioms 5, 6 and 7.*

Such a result is thus equivalent to the seminal characterization of the *Shapley value* for TU-games (e.g., Shapley, 1953).[14] The Shapley value is the best-known solution concept for those games, which model the attempt to predict the allocation of resources in multi-person interactions (see Winter, 2002). It is remarkable not only for its attractive and intuitive definition but also for its unique characterization by a set of reasonable axioms. In addition, the value is also viewed as an index for measuring the power of players (here, languages) in a game (see Shapley and Shubik, 1954). The value uses averages (or weighted averages in some of its generalizations) to aggregate the power of players in their various cooperation opportunities.

One of the axioms characterizing the Shapley value is *Equal Treatment of Equals*. This is a compelling axiom when information is limited. For some specific applications, however, we might possess more information about the environment, which motivates the asymmetry of the solution concept.[15] This interpretation led to the concept of *Weighted Shapley Value*, which has also been characterized in the literature, replacing the axiom *Equal Treatment of Equals* by *Partnership Consistency*, which refers to the treatment of players who can only generate value together (see Kalai and Samet, 1987). We consider another related alternative to *Equal Treatment of Equals*, which leads to the characterization of WS.

Axiom 5' Weighted Treatment of Equals. If a speaker speaks two languages, then she should allocate vote shares among them that are proportional to their knowledge levels.

This leads to the characterization of the WS ranking:

Proposition 5. *WS is the unique ranking that satisfies Axioms 5', 6 and 7.*[16]

[14] A coalitional game is cooperative if the players can make binding agreements about the distribution of payoffs or the choice of strategies, even if these agreements are not specified or implied by the rules of the game. Transferable-utility games (in short, TU-games) are one category of cooperative games in which one specifies a function that associates with each nonempty coalition a real number indicating the worth of the coalition. If a coalition forms, then it can divide its worth in any possible way among its members. A comprehensive analysis of TU-games can be found, for instance, in Peleg and Sudholter (2007).

[15] In our context, this is so if we allow for different knowledge levels of languages.

[16] See Appendix 1 for the formal proof.

3 Languages in the European Union

We now use the framework described in Section 2 to rank the main languages used in the EU. We distinguish official languages (see below) from other languages that are spoken, but are not official.

Official languages. In 1958, the Treaty of Rome and Regulation 1/1958 recognized four languages Dutch (NL), French (F), German (G) and Italian (I) as official languages.[17] Danish (DK), English (GB), Finnish (FIN), Greek (GR), Portuguese (P), Spanish (SP) and Swedish (SW) were added later. The 2004 enlargement to Eastern Europe resulted in adding Czech (CZ), Estonian (EST), Hungarian (H), Latvian (LV), Lithuanian (LT), Maltese (M), Polish (PL), Slovak (SL), and Slovenian (SLO). Irish (IRL) was given the same status in 2005 but it was agreed that the decision would be implemented only as of January 2007. We also included Bulgarian (BG) and Romanian (RO) which became official in 2007 only, but were already included in the survey to be discussed below.[18] All these languages, listed as *Official* in Table 1 enjoy the same privileges as the original four.

[Table 1 approximately here]

Other languages. Table 1 also includes seven other languages that are used in EU countries, but are not official. Russian is spoken in many former Eastern Bloc countries; Basque, Catalan and Galician are spoken in Spain, and Luxembourgish in Luxembourg; Arabic and Turkish are languages mainly spoken by immigrants.[19]

To determine who speaks what, we use the Special Eurobarometer 243 (2006) survey carried out in November 2005 in all member countries of the EU, including Bulgaria and Romania (that were not yet members in 2005). In most countries, 1,000 citizens were interviewed, with the exception of Germany (1,500), the United Kingdom (1,300),

[17] Though the Treaty included six countries, the three languages that are spoken in Belgium (French, Dutch and German) are also spoken in other partner countries (France, the Netherlands and Germany); Luxembourg agreed not to include Luxembourgish (LX).

[18] Croatia and Turkey were also candidates to accessing the EU, but this happened much later for Croatia, and did not happen for Turkey. Croatian is thus not included in our list. Turkish is, but only as a non-official language, spoken by migrants, and as a co-official language in Cyprus.

[19] Turkish is also, as mentioned above, a co-official language in Cyprus, which is a country included in our survey.

Cyprus (500), Luxembourg (500) and Malta (500). The total number of usable interviews amounts to 26,700.[20]

The data that we use are taken from the answers to the following questions:

(a) D48a. What is your mother tongue? (do not probe – do not read out – multiple answers possible). Follows a list of 34 languages that include the 23 member states' official languages, as well as Arabic, Catalan, Chinese, Croatian, Luxembourgish, Russian, Turkish, Basque, Galician, Other regional languages, Other.

(b) D48b to D48d. Which languages do you speak well enough in order to be able to have a conversation, excluding your mother tongue? (do note probe – do not read out – multiple answers possible). Follows the same list of 34 languages. This question was asked for first, second and third foreign languages.

(c) D48f. Is your (language cited in 48b, 48c and 48d) very good, good, basic? (show card with scale).

In our calculations, we code as 1 the responses *mother tongue* (question D48a), and as 1/2 the responses to other languages known (questions D48b to D48d), combined with the self-evaluation of knowledge *very good* and *good* (question D48f). All other responses, that is *basic*, or *not knowing the language* are coded as 0.

Table 1 summarizes our findings. Note that we separate Russian, Arabic, Turkish, Catalan, Basque, Galician as well as Luxembourgish, which appear at the bottom of the table, as they are not official.[21] These results are obtained by considering the whole EU as

[20]It is worth mentioning the following technical specifications which ensure representativity:

"[...] the survey covers the national population of citizens [...] that are residents in those countries and have a sufficient command of one of the respective national language(s) to answer the questionnaire. The basic sample design applied in all states is a multi-stage, random (probability) one. In each country, a number of sampling points was drawn with probability proportional to population size (for a total coverage of the country) and to population density.

For each country a comparison between the sample and the universe was carried out. The universe description was derived from Eurostat population data or from national statistics offices. For all countries surveyed, a national weighting procedure, using marginal and intercellular weighting, was carried out based on this universe description. In all countries, gender, age, region and size of locality were introduced in the iteration procedure."

[21]With the caveat made before about Turkish in Cyprus.

a unique country. As the number of units surveyed are not proportional to the populations of the 27 countries (Croatia is not included, see footnote 16) included, we had to weigh the numbers of each country by its population.[22]

In the table, column (1) contains the name of the language, and column (2) the countries whether it is official. Columns (3) to (5) contain the results of the first three rankings (*Minimal Disenfranchisement* (MD), *Communicative Benefits* (CB), and *Aggregate Knowledge* (AK)). The numbers represent the shares of the total EU population (490 million at the time of the survey) that do not know the language, according to the chosen criterion. The rankings can thus be easily interpreted in terms of numbers of citizens. Columns (6) and (7) are devoted to the *Shapley* ranking (S) and the *Weighted Shapley* ranking (WS). The interpretation is less obvious, but the distances between languages can be interesting to consider. Both types of rankings include therefore, to some extent, a cardinal aspect, which indicate big and small jumps between languages:[23]

(a) column (3) (MD) contains the shares of the total EU population that does not know the language or whose knowledge is basic; for instance, the value 63 that appears in the first row of column (3), means that 63 percent of the EU population does not know or only has a basic knowledge of English.[24]

(b) column (4) (CB): the numbers represent the shares of the total EU population which do not speak the language as mother tongue; the value 87, which appears in the first row of column (4), means that 87 percent of the EU population do not have English as mother tongue.

(c) column (5) (AK): the numbers are less obvious to be interpreted in a simple way, as they result from adding knowledge levels (a_{ij}) of which some are equal to 1, others are equal to 1/2. For consistency, we adopted the same convention as in the previous two rankings and express the numbers in terms of the percentages of the EU population. For example, the value 75, which appears in the first row of column

[22]More precisely, the population (in millions) we used for each country was the following: Austria 8.2; Belgium 10.4; Bulgaria 7.8; Cyprus 0.7; Czech Rep. 10.2; Denmark 5.4; Estonia 1.3; Finland 5.2; France 60.6; Germany 82.5; Great Britain 60; Greece 11.1; Hungaria 10.1; Ireland 4.1; Italy 58.5; Latvia 2.3; Lithuania 3.4; Luxembourg 0.5; Malta 0.4; Netherlands 16.3; Poland 38.2; Portugal 10.5; Romania 21.7; Slovakia 5.4; Slovenia 2; Spain, 43; Sweden 9.

[23]The numbers that appear in column (3) guided the ordering in which the languages appear in the table. This is of course arbitrary, as any other column could have been chosen as well.

[24]The 100 numbers appearing in this column, as well as the next ones, are a consequence of rounding, which we did to the closest integer percentages.

(5), means that 25 percent of the EU population is obtained when we aggregate all individuals having English as mother tongue and one half of all individuals having good (or very good) knowledge of English.

(d) (6) and (7) (S) and (WS): the rankings using *Shapley Values* have no meaning other than giving a measure of the power of a language; what really counts here is the order, and to some extent, the differences between languages. One could surmise that English and Swedish are far away from each other, but there is no difference between Swedish and Bulgarian.

The main lesson to be drawn from our empirical analysis is that, in spite of the theoretical differences and distinct normative grounds, the rankings methods we consider yield consistent results. The *Communicative Benefits* ranking, in column (4), points, nevertheless, to a couple of important differences with respect to the *Minimal Disenfranchisement* ranking in column (3). Ranking CB only considers native speakers, while ranking AK contains native speakers (who are counted as 1) and those who speak the language very well or well (who are counted as 1/2). Note that only three languages are sensitive to the changes: English, German and French. The main shock is on English, since the English population (60 million) is smaller than the German and Austrian populations (91 million), and many more people speak English well (though not as a native language) so that English gains a lot going from CB to AK (or MD), because many English speakers are added by AK (or MD). Most languages in the first half part of the ranking disenfranchise more citizens than in column (3), as they are not counted if not perfectly spoken. German is the most obvious change since it is the language with the highest number of native speakers, but it also is a language that less non-native individuals speak (even at intermediate levels), compared to English or French. English, French and Italian would generate identical levels of disenfranchisement.

In the *Shapley* ranking, the downgrading of Dutch is due to the fact that the number of Netherlanders and Flemish (in Belgium) who speak foreign languages (English, German and French essentially) is large, which reduces the Shapley Value of Dutch. The upgrading of Romanian is the consequence of the inverse phenomenon: Most of Romanians are unilingual, which increases the Shapley Value. In the (WS) ranking, Romanian raises to rank 5, and comes before Spanish and Polish.

Russian is an interesting case as it is one of the ten most important languages in the EU (with the exception of the CB ranking). This is, of course, due to the Russian presence in the Baltic republics and its influence in Eastern Europe after Word War II, where it is

often a second language only (which explains its relatively low standing in the CB, S and WS ranking). The forces in action nowadays may well reduce the extent of knowledge of Russian among the younger generations, but the expansion of trade and business ties with Russia may foster the learning of Russian.

4 The consequences of Brexit

English is not only a so-called *official* language of the EU, but it is also the one that is most widely spoken as native (or well-known) by some 182 million out of 490 millions Europeans. It became de facto the *lingua franca* in Europe, though it disenfranchises large groups in many countries.[25] Following the pro-Brexit referendum held in the UK in June 2016, two main speculative scenarios are possible.[26] Each of them leads to conclusions that are also speculative.

One scenario is that, after Brexit occurs, English loses its official status in the EU,[27] though it is an official language in Ireland (together with Gaelic) and in Malta (together with Maltese). It is indeed spoken by a very large majority of citizens in Ireland, which accessed the EU in 1973. Some 25 years later, it asked Gaelic to become its official language. This was accepted by the EU, and on January 1, 2007 Gaelic became one of the 24 official languages, with some derogations, that should however be brought to an end by December 31, 2021.[28] English is also spoken by a majority of Maltese. Accepting that Ireland (4.1 million inhabitants) and Malta (0.4 million inhabitants) each have two official languages (though English is common) may trigger other countries (or regions) to get their language accepted. Catalan that is spoken by 6.2 million people in Catalonia may be first in a row. Galician may follow suit with 3 million speakers, as well as Basque with 0.7 million people. All three languages are spoken by more people than Maltese and

[25] See Fidrmuc et al.(2007).

[26] We assume that Brexit will indeed happen sometime in 2019. See, for instance, http://www.bbc.com/news/uk-politics-37532364, consulted on November 28, 2016.

[27] Danuta Hübner, head of the European Parliament's Constitutional Affairs Committee (AFCO) warned on June 26 that English will lose its official status. See Hortense Goulard, English will not be an official EU language after Brexit, says senior MEP. See http://www.politico.eu/article/english-will-not-be-an-official-eu-language-after-brexit-senior-mep/, consulted on August 16, 2016. This has been countered by the Commission's representation in Ireland on June 28. See https://www.euractiv.com/section/languages-culture/news/english-will-remain-an-official-language-of-the-eu/, consulted on November 3, 2016.

[28] See Europa, Interinstitutional style guide, section 7.2.4, Rules governing the languages of the institutions.

Gaelic, and Catalan is spoken by more people than the whole population of Ireland. It may be that either Ireland or Malta accept to replace their official language by English, but it may also be politically difficult for both of them to face their own populations with such a proposal.

The other scenario is that the EU considers nevertheless keeping English as an official language. This may be rebuked by some countries, and if so, require a vote by the European Council.[29] According to the EU's constitutional rules, votes on languages have to be unanimous. It is doubtful that Germany and France will cast a positive vote, as they may prefer their official languages replace English as the largely used working language.[30]

We now explore the two scenarios using three of the rankings discussed earlier (*Minimal Disenfranchisement*, *Communicative Benefits* and *Aggregate Knowledge*). We start with some intuitive numbers, and will refine this later using rankings. These numbers are not influenced by the fact that English may lose its official status in the EU. They are just the consequence of the UK leaving the EU.

In 2005, the EU had some 490 million citizens. 182.5 of them were reasonably fluent in English. German (128.5 million), French (100.7 million), Italian (68.4 million), Spanish (57.2 million) and Polish (43.5 million) followed.[31] Brexit has two effects. The important one is the loss of some 60 million British native speakers, so that the 182.5 millions will drop to 122.5 million. The second effect, which is much less important, is concerned with the reduction of Britons and immigrants who live in the UK, and who know one of the other EU languages, in particular some 2 and 6 millions speakers of French and German. Other languages will also incur some losses, though these are smaller (Italian, -1 million, Spanish, -1.3 million).

[Table 2 approximately here]

This leads us to columns MD in Table 2, where we compare *Minimal Disenfranchisement* (which includes native speakers of a language as well as those who speak it very

[29]Unless Scotland secedes from the post-Brexit UK, and it is accepted as an EU member with English as its official language.

[30]English is also heavily used in reports and rules set by EU bureaucrats. In 2008, 72.5 percent of the first versions of administrative and legislative documents were written in English, while 14 and 3 percent only were written in French and German, respectively. The situation has even become more skewed as English now makes for some 82.5 percent, while German dropped to 2 percent. To top things off, in 2014, 261.000 pages have been translated from other languages into English, 155.000 into French and 136.000 into German. This means that the EU essentially writes in English and translates into English.

[31]These numbers include native speakers, as well as non natives who acquired the language, and know it very well or well.

well and well) before (column B) and after (column A) Brexit, taking into account that the number of EU citizens dropped from 488 to 428 millions, as 60 million Britons will have left the EU. As can be seen, the number of speakers of German and English are equal after Brexit, which was far from being the case before. There is still a 'qualitative' difference, as those who speak English (121 million) are not living in the UK, while 91 million among the 121 million speakers of German are living in German speaking countries (Austria and Germany). Therefore, English could still keep its *lingua franca* influence in the EU, whether or not it loses its official status.

The CB (*Communicative Benefits*) rankings in Table 2 are based on native speakers only. Here, the problem of English is getting very acute, as it will be spoken by some 6.5 million only, of which 4.1 live in Ireland. So English does not do better than any language that is spoken by more than 6.5 millions inhabitants, which includes Bulgarian, Czech, Dutch, French, German, Greek, Hungarian, Italian, Polish, Portuguese, Romanian, Spanish and Swedish, that is, more than half of EU's official languages.

The *AK* (*Aggregate Knowledge*) ranking, in which native speakers are counted for one, while others, who speak it very well or well are counted for 1/2, produces a ranking in which German (76) and French (82) are better off than English (85), while Italian (86) and Spanish (89) are very close to English in terms of communication, though all of them are less widespread across the EU than English.

In conclusion, English remains a powerful language in Europe even after *Brexit*. The real question is whether it will remain an official language in the EU. If so, it will be interesting to see what kind of arguments will be used by the European Commission. Making it an official language in Ireland and Malta might not be accepted by the Germans and the French, and/or it opens the possibility of other languages becoming official (Catalan, in particular). If English loses its official status in the EU, it will no longer be used in the European Parliament nor will official documents be translated into English. The question is whether it will be able to maintain its status as working language. The odds are good, as it is one of the important relay languages used by interpreters in the EU Parliament, and probably also by those who translate official documents. Using another relay language would certainly dramatically increase, at least during the transition (which could take years), costs and create disruptions of the entire translation process.

5 Conclusions

In this paper, we provide a stylized framework for ranking languages in multilingual societies. We introduce five ranking methods. Three of these reflect appealing principles such as minimizing exclusion (MD), maximizing communication benefits (CB), or aggregating knowledge levels (AK). The two last, (S) and (WS), are inspired by game-theoretical concepts. We explore the normative foundations for each method and apply them for ranking languages in the EU. The reader may ask why we tried several rankings. The main reason is that each ranking is based on different axioms, and it was interesting to see to which results they were leveling. As it can be seen, the results are remarkably similar in our empirical application. The reader may also ask which ranking is believed to be the best. We would favor MD and AK, which are identical (at least before Brexit). They can also be interpreted in terms of the share of the population that will be disenfranchised if the language were chosen as official, or even working, language. The Shapley rankings are less intuitive, though they are not very different from the previous ones in our empirical application. If used, they may favor populations that know few languages.

One should add that people learn new languages, but may also stop learning languages which become less important de facto, or in their minds. It could happen that, in the after-Brexit world, EU citizens may decide to forego speaking or learning English. Moreover, as suggested by Ginsburgh et al. (2016) English has no particular advantage over other languages, in particular if trade with China, India, Brazil, Africa or the Arab world becomes more important. This implies that the rankings computed in this paper may be prone to change over time.

6 References

Alcalde-Unzu, J., and Vorsatz, M. (2009) Size approval voting. Journal of Economic Theory 144, 1187-1210.

Alesina, A., and La Ferrara, E. (2005) Ethnic Diversity Economic Performance. Journal of Economic Literature 43, 762-800.

Athanasiou, E., Moreno-Ternero, J.D., and Weber, S. (2016) Language learning and communicative benefits. In V. Ginsburgh and S. Weber (eds.), The Palgrave Handbook of Economics and Language, London: Palgrave Macmillan, 212-229.

Bergantiños, G., and Moreno-Ternero, J.D. (2015) The axiomatic approach to the problem of sharing the revenue from museum passes. Games and Economic Behavior 89, 78-92.

Blackorby, C., Bossert, W., and Donaldson, D. (2002) Utilitarianism and the theory of justice,. In K. J. Arrow, A. Sen, and K. Suzumura (eds.), The Handbook of Social Choice and Welfare, Amsterdam: Elsevier, 543-596.

Brams, S., and Fishburn, P. (1978) Approval voting. American Political Science Review 72, 831-847.

Europa International institutional style guide. http://publications.europa.eu/code/en/en-370204.htm.

Fidrmuc, J., Ginsburgh, V. and Weber, S. (2007), Ever Closer Union or Babylonian Discord? The Official-language Problem in the European Union, CEPR Discussion Paper 6397.

Ginsburgh, V., and Weber, S. (2005) Language disenfranchisement in the European Union. Journal of Common Market Studies 43, 273-286.

Ginsburgh, V., and Weber, S. (2011), *How Many Languages Do We Need? The Economics of Linguistic Diversity*, Princeton, NJ: Princeton University Press.

Ginsburgh, V., Melitz, J., and Toubal, F. (2016) Foreign language learning and trade, Review of International Economics, forthcoming.

Ginsburgh, V., Ortuño-Ortin, I., and Weber S. (2005) Language disenfranchisement in linguistically diverse societies. The case of European Union. Journal of the European Economic Association 3, 946-965.

Ginsburgh, V., and Zang, I. (2003) The museum pass game and its value. Games and Economic Behavior 43, 322-325.

Ginsburgh, V., and Zang, I. (2012) Shapley ranking of wines. Journal of Wine Economics 7, 169-180.

Glasser, G. (1959) Game theory and cumulative voting for corporate directors. Management Science 5, 151-156.

Hillinger, C. (2005) The case for utilitarian voting. Homo Oeconomicus 23, 295-321.

Holden, N. (2016) Economic exchange and business languages in the ancient world: an exploratory review. In V. Ginsburgh and S. Weber (eds.), The Palgrave Handbook of Economics and Language, London: Palgrave Macmillan, 290-311.

Kadochnikov, D. (2016) Languages, regional conflicts and economic development: Russia. In V. Ginsburgh and S. Weber (eds.), The Palgrave Handbook of Economics and Language, London: Palgrave Macmillan, 538-580.

Kalai, E., and Samet, D. (1987) On Weighted Shapley Values. International Journal of Game Theory 16, 205-222.

Macé, A. (2015) Voting with evaluations: When should we sum? What should we sum? AMSE Working Paper

Van Parijs, P. (2011) Linguistic Justice for Europe and for the World. Oxford: Oxford University Press.

Peleg, B., and Suholter, P. (2007) Introduction to the Theory of Cooperative Games. Springer

Sawyer, J., and MacRae, D. (1962) Game theory and cumulative voting in Illinois: 1902-1954. American Political Science Review 56, 936-946.

Selten, R., and Pool, J. (1991) The distribution of foreign language skills as a game equilibrium. In Selten, R. (ed.). Game Equilibrium Models, Berlin: Springer-Verlag, 64-84.

Shapley, L. (1953) A value for n-person games. In H.W. Kuhn and A.W. Tucker (eds.), Contributions to the Theory of Games II (Annals of Mathematics Studies 28), Princeton: Princeton University Press, 307-317.

Shapley, L., and Shubik, M. (1954) A method for evaluating the distribution of power in a committee system. American Political Science Review 48, 787-792.

Sonntag, S. (2016) Languages, regional conflicts and economic development in South Asia. In V. Ginsburgh and S. Weber (eds.), The Palgrave Handbook of Economics and Language, London: Palgrave Macmillan, 489-512.

Special Eurobarometer 243, 2006.

Winter, E. (2002) The Shapley value. In R. Aumann, and S. Hart (eds.) Handbook of Game Theory, Volume 3. Elsevier, pp 2025-2054.

Wright, S. (2016) Language Policy and Language Planning. London: Palgrave MacMillan.

Xu, Y. (2010) Axiomatizations of approval voting. In J.-F. Laslier, and R. Sanver (eds.), Handbook on Approval Voting. Studies in Choice and Welfare, Berlin Heidelberg: Springer-Verlag, 91-102

7 Appendix

We provide in this appendix a formal treatment to the contents of Section 2.2.

For each language matrix A, we construct two matrices associated to it: A^L and A^H. Formally, for each $(i,j) \in N \times L$,

$$a_{ij}^L = \begin{cases} 0 & \text{if } a_{ij} < 1 \\ 1 & \text{otherwise} \end{cases}$$

and

$$a_{ij}^H = \begin{cases} 1 & \text{if } a_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $\succsim_A$ denote the ranking of languages associated to A. Our axioms are formally stated as follows:

- *Low Invariance*: For each pair of languages $l, k \in L$, $l \succsim_A k \iff l \succsim_{A^L} k$.
- *High Invariance*: For each pair of languages $l, k \in L$, $l \succsim_A k \iff l \succsim_{A^H} k$.
- *Impartiality*: Let $l, k \in L$ and $i, j \in N$ be such that $a_{il} = 1 = a_{jk}$, $a_{jl} = 0 = a_{ik}$, and $a_{hl} = a_{hk}$, for all $h \in N \setminus \{i, j\}$. Then, $l \sim_A k$.
- *Strong Pareto*: Let $l, k \in L$ be such that $a_{il} \geq a_{ik}$, for each $i \in N$, with at least one strict inequality. Then, $l \succ_A k$.

The formal definition of our rankings comes next.

- *Minimal Disenfranchisement*: For each pair of languages $l, k \in L$,

$$l \succsim_A^{MD} k \iff MD(l) = \#\{i \in N | a_{il} = 0\} \leq MD(k) = \#\{i \in N | a_{ik} = 0\}$$

- *Communicative Benefits*: For each pair of languages $l, k \in L$,

$$l \succsim_A^{CB} k \iff CB(l) = \#\{i \in N | a_{il} = 1\} \geq CB(k) = \#\{i \in N | a_{ik} = 1\}$$

- *Aggregate Knowledge*: For each pair of languages $l, k \in L$,

$$l \succsim_A^{AK} k \iff AK(l) = \sum_{i \in N} a_{il} \geq AK(k) = \sum_{i \in N} a_{kl}$$

We have the following characterizations:

Proposition 1 *Minimal Disenfranchisement is the unique ranking satisfying High Invariance, Impartiality and Strong Pareto.*

Proof. Let $A = (a_{ij})_{(i,j)\in N\times L}$ be a language matrix, and $l, k \in L$ be a pair of languages. By *High Invariance*, $l \succsim_A k \iff l \succsim_{A^H} k$. By iterated application of *Impartiality*, if necessary, we obtain that $l \sim_{A^H} k \iff \sum_{i\in N} a^H_{il} = \sum_{i\in N} a^H_{ik}$. By *Strong Pareto*, $l \succ_{A^H} k \iff \sum_{i\in N} a^H_{il} > \sum_{i\in N} a^H_{ik}$. Altogether, we have that $l \succsim_A k \iff MD(l) \leq MD(k)$. ■

Note that *Communicative Benefits* is a ranking satisfying *Impartiality* and *Strong Pareto* but not *High Invariance*. The Shapley ranking satisfies *High Invariance* and *Strong Pareto* but not *Impartiality*. Finally, the *Anti-Minimal Disenfranchisement* is a ranking satisfying *High Invariance* and *Impartiality* but not *Strong Pareto*.[32] Thus, all axioms in Proposition 1 are independent.

Proposition 2 *Communicative Benefits is the unique ranking satisfying Low Invariance, Impartiality and Strong Pareto.*

Proof. Let $A = (a_{ij})_{(i,j)\in N\times L}$ be a language matrix, and $l, k \in L$ be a pair of languages. By *Low Invariance*, $l \succsim_A k \iff l \succsim_{A^L} k$. By iterated application of *Impartiality*, if necessary, we obtain that $l \sim_{A^L} k \iff \sum_{i\in N} a^L_{il} = \sum_{i\in N} a^L_{ik}$. By *Strong Pareto*, $l \succ_{A^L} k \iff \sum_{i\in N} a^L_{il} > \sum_{i\in N} a^L_{ik}$. Altogether, we have that $l \succsim_A k \iff CB(l) \geq CB(k)$ ■

Note that *Minimal Disenfranchisement* is a ranking satisfying *Impartiality* and *Strong Pareto* but not *Low Invariance*; the *Perfect Shapley* ranking satisfies *Low Invariance* and *Strong Pareto* but not *Impartiality*;[33] and the *Anti-Communicative Benefits* ranking satisfies *Low Invariance* and *Impartiality* but not *Strong Pareto*.[34] Thus, all axioms in Proposition 2 are independent.

[32]The *Anti-Minimal Disenfranchisement* ranking is the one that ranks languages exactly in the opposite way as the *Minimal Disenfranchisement* ranking. Formally, according to the *Anti-Minimal Disenfranchisement* ranking, language j is ranked higher than language k if $\#\{i \in N | a_{ij} = 0\} > \#\{i \in N | a_{ik} = 0\}$.

[33]The *Perfect Shapley* ranking is the one that replaces in the Shapley ranking languages with some knowledge by languages with perfect knowledge. Formally, let $P_i(A) = \{j \in L : a_{ij} = 1\}$ denote the set of languages agent $i \in N$ has perfect knowledge of, and let $p_i(A)$ denote the number of such languages. Then, we say that language j is ranked higher than k, according to the *Perfect Shapley* ranking, if

$$\sum_{i\in N, j\in P_i(A)} \frac{1}{p_i(A)} > \sum_{i\in N, k\in P_i(A)} \frac{1}{p_i(A)}.$$

[34]The *Anti-Communicative Benefits* ranking is the one that ranks languages exactly in the opposite way as the *Communicative Benefits* ranking. Formally, according to the *Anti-Communicative Benefits* ranking, language j is ranked higher than language k if $\#\{i \in N | a_{ij} = 1\} < \#\{i \in N | a_{ik} = 1\}$.

In order to define the next axiom, suppose now that $a_{ij} \in \{a_1, a_2, \ldots, a_n\}$, where $a_1 < a_2 < \cdots < a_n$.

- *Generalized Impartiality*: Let $l, k \in L$ and $i, j \in N$ be such that $a_{il} = a_m = a_{jk}$, $a_{jl} = a_{m-1} = a_{ik}$, for some $m \in \{2, \ldots, n\}$, and $a_{hl} = a_{hk}$, for all $h \in N \setminus \{i, j\}$. Then, $l \sim_A k$.

Proposition 3 *Aggregate Knowledge is the unique ranking satisfying Generalized Impartiality and Strong Pareto.*

Proof. Let $A = (a_{ij})_{(i,j)\in N\times L}$ be a language matrix, and $l, k \in L$ be a pair of languages. By iterated application of *Generalized Impartiality*, if necessary, we obtain that $l \sim_A k \iff \sum_{i\in N} a_{il} = \sum_{i\in N} a_{ik}$. By *Strong Pareto* and *Generalized Impartiality*, $l \succ_A k \iff \sum_{i\in N} a_{il} > \sum_{i\in N} a_{ik}$. Altogether, $l \succsim_A k \iff AK(l) \geq AK(k)$.[35] ■

Note that the *Minimal Disenfranchisement* ranking satisfies *Strong Pareto* but not *Generalized Impartiality* and the *Anti-Aggregate Knowledge* ranking satisfies *Generalized Impartiality* but not *Strong Pareto*.[36] Thus, all axioms in Proposition 3 are independent.

Suppose now that each agent has a *vote* to be allocated among the languages she speaks. In this context, a **rule** is a mapping R that associates with each language matrix A an allocation $(R_l(A))_{l\in L}$ indicating the amount of votes each language gets. The Shapley rule allocates for each language $l \in L$ the following votes:

$$R_l^S(A) = \sum_{i\in N, l\in L_i(A)} \frac{1}{l_i(A)},$$

where recall that $L_i(A) = \{j \in L : a_{ij} \neq 0\}$ and $l_i(A)$ denotes its cardinality.

This rule obviously leads to the Shapley ranking we introduced above. As shown by the next result, which replicates Theorem 1 in Bergantiños and Moreno-Ternero (2015), the Shapley rule is characterized by the following three axioms:

- *Equal treatment of equals*: Let $l, k \in L$ be such that $N_l(A) = \{i \in N : a_{il} \neq 0\} = \{i \in N : a_{ik} \neq 0\} = N_k(A)$. Then, $R_l(A) = R_k(A)$.

[35] This proof essentially mimicks the proof of Theorem 3 in Macé (2015).

[36] The *Anti-Aggregate Knowledge* ranking is the one that ranks languages exactly in the opposite way as the *Aggregate Knowledge* ranking. Formally, according to the *Anti-Aggregate Knowledge* ranking, language j is ranked higher than language k if $\sum_{i\in N} a_{ij} < \sum_{i\in N} a_{ik}$.

- *Dummy*: Let $l \in L$ be such that $N_l(A) = \emptyset$. Then, $R_l(A) = 0$.
- *Additivity*: For each pair of subsets of agents (N^1, N^2), and their corresponding language matrices A^1 and A^2, $R(A^1 \cup A^2) = R(A^1) + R(A^2)$, where $A^1 \cup A^2$ denotes the resulting matrix from merging the rows of both matrices.

Proposition 4 *A rule satisfies equal treatment of equals, dummy and additivity if and only if it is the Shapley rule.*

Proof. It is obvious that the Shapley rule satisfies the axioms. Let R be a rule satisfying the three axioms in the statement. For each $i \in N$, let A_i denote the corresponding i-th row of A. By *dummy*, $R_l(A_i) = 0$ for each $l \notin L_i(A)$. By *equal treatment of equals*, $R_l(A_i) = R_k(A_i)$ for each pair $l, k \in L_i(A)$. As $\sum_{l \in L} R_l(A_i) = 1$, we deduce that $R_l(A_i) = \frac{1}{l_i(A_i)}$ for each $l \in L_i(A)$. Note that $l_i(A_i) = l_i(A)$. Consequently, it follows, by *additivity*, that $R_l(A) = \sum_{i \in N, j \in L_i(A)} \frac{1}{l_i(A)}$, for each $l \in L$, as desired. ■

Note that Bergantiños and Moreno-Ternero (2015) show that all axioms in Proposition 4 are independent.

To conclude, we consider the Weighted Shapley rule, which allocates for each language $l \in L$ the following votes:

$$R_l^{WS}(A) = \sum_{i \in N, l \in L} \frac{a_{il}}{\sum_{h \in L} a_{ih}}.$$

As shown in the last proposition, the following alternative axiom to *equal treatment of equals* leads to characterize the Weighted Shapley rule.

- *Weighted treatment of equals*: Let $i \in N$ and $l, k \in L_i(A)$. Let A_i denote the corresponding i-th row of A. Then, $\frac{R_l(A_i)}{R_k(A_i)} = \frac{a_{il}}{a_{ik}}$.

Proposition 5 *A rule satisfies weighted treatment of equals, dummy and additivity if and only if it is the Weighted Shapley rule.*

Proof. It is obvious that the Weighted Shapley rule satisfies the axioms. Let R be a rule satisfying the three axioms in the statement. For each $i \in N$, let A_i denote the corresponding i-th row of A. By *dummy*, $R_l(A_i) = 0$ for each $l \notin L_i(A)$. By *weighted treatment of equals* $\frac{R_l(A_i)}{R_k(A_i)} = \frac{a_{il}}{a_{ik}}$ for each pair $l, k \in L_i(A)$. As $\sum_{l \in L} R_l(A_i) = 1$, we deduce that $R_l(A_i) = \frac{a_{il}}{\sum_{h \in L} a_{ih}}$ for each $l \in L_i(A)$. Consequently, it follows, by *additivity*, that $R_l(A) = \sum_{i \in N, l \in L} \frac{a_{il}}{\sum_{h \in L} a_{ih}}$, for each $l \in L$, as desired. ■

Note that the examples considered by Bergantiños and Moreno-Ternero (2015), to show that the axioms in Proposition 4 are independent, can easily be adjusted to prove that so are the axioms in Proposition 5.

Table 1. Rankings of Languages in the EU Before Brexit

Languages		Rankings				
Name (1)	Native in (2)	MD (3)	CB (4)	AK (5)	S (6)	WS (7)
Official						
English	GB-IRL	63	87	75	106	96
German	G-A-B	75	82	79	78	54
French	F-B	80	87	84	64	43
Italian	I	87	88	87	49	35
Spanish	E	89	92	90	38	26
Polish	PL	92	92	92	32	22
Dutch	NL-B	95	96	95	12	10
Romanian	RO	95	96	96	19	28
Hungarian	H	97	98	97	11	8
Greek	GR-CY	97	98	97	10	7
Portuguese	P	98	98	98	10	7
Czech	CZ	98	98	98	8	6
Swedish	S	98	98	98	6	5
Bulgarian	BG	98	99	98	6	4
Slovak	SL	99	99	99	4	3
Danish	DK	99	99	99	3	3
Finnish	FIN	99	99	99	3	3
Lituanian	LT	99	99	99	2	2
Slovenian	SLO	99	100	99	1	1
Latvian	LV	100	100	100	1	1
Estonian	EST	100	100	100	<1	<1
Irish	IRL	100	100	100	<1	<1
Maltese	M	100	100	100	<1	<1
Others						
Russian	Non EU	95	99	97	10	6
Catalan	SP	99	99	99	2	2
Galician	SP	99	100	99	1	1
Arabic	Non EU	99	100	100	1	1
Turkish	CY	100	100	100	1	1
Basque	SP	100	100	100	<1	<1
Luxemb.	LX	100	100	100	<1	<1

Table 2. Rankings of Main Languages in the EU
Before (B) and After (A) Brexit

Languages		Rankings					
Name	Native in	MD		CB		AK	
(1)	(2)	B	A	B	A	B	A
English	GB-IRL	63	72	87	99	75	85
German	G-A-B	75	72	82	80	79	76
French	F-B	80	79	87	85	84	82
Italian	I	87	85	88	87	87	86
Spanish	E	89	88	92	91	90	89
Polish	PL	92	91	92	91	92	91
Dutch	NL-B	95	94	96	95	95	95
Romanian	RO	95	95	96	95	96	95
Hungarian	H	97	97	98	97	97	97
Greek	GR-CY	97	97	98	97	97	97
Portuguese	P	98	97	98	97	98	97
Czech	CZ	98	97	98	98	98	97
Swedish	S	98	98	98	98	98	98
Bulgarian	BG	98	98	99	98	98	98
Slovak	SL	99	98	99	99	99	99
Danish	DK	99	99	99	99	99	99
Finnish	FIN	99	99	99	99	99	99
Lituanian	LT	99	99	99	99	99	99
Slovenian	SLO	99	99	100	99	99	99
Latvian	LV	100	100	100	100	100	100
Estonian	EST	100	100	100	100	100	100
Irish	IRL	100	100	100	100	100	100
Maltese	M	100	100	100	100	100	100
Others							
Russian	Non EU	95	95	99	99	97	97
Catalan	SP	99	99	99	99	99	99
Galician	SP	99	99	100	100	99	99
Arabic	Non EU	99	100	100	100	100	100
Turkish	CY	100	99	100	100	100	100
Basque	SP	100	100	100	100	100	100
Luxemb.	LX	100	100	100	100	100	100