The "own-language preference": measures of "relative language self-citation"

Peer-reviewed author version

EGGHE, Leo; ROUSSEAU, Ronald & Yitzhaki, M. (1999) The "own-language preference": measures of "relative language self-citation". In: Scientometrics, 45(2). p. 217-232.

DOI: 10.1007/BF02458434
Handle: http://hdl.handle.net/1942/790

THE 'OWN-LANGUAGE PREFERENCE': MEASURES OF `RELATIVE LANGUAGE SELF-CITATION'

Leo Egghe*, Ronald Rousseau* and Moshe Yitzhaki**

* LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
and UIA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium

** Bar-Ilan University, Dept. Information and Library Studies
Ramat-Gan, 52900 Israel

Abstract

It has already been pointed out that the foreign language barrier is probably the greatest impediment to the free flow and transfer of information. This barrier is even growing as scientists of more and more countries publish in their own languages. Almost all studies addressing the language barrier problem were conducted from an Anglo-Saxon perspective, limiting their scope to English-language sources or English speakers. Little research has been devoted to studying and measuring language preference among non-English-speaking scholars.

This article reviews measures proposed in former studies such as the 'relative own-language preference' indicator, and the 'straight odds ratio', pointing out their advantages and drawbacks. Two new refined measures (in both 'raw' and normalised versions) are offered, claiming to be free of these drawbacks, and thus enabling a better and more reliable comparison between journals of different languages. Practical use of the proposed measures is illustrated by applying them to findings of a former language-citation study done on nine sociology journals.

Introduction

Large[1] already pointed out that the foreign language barrier is probably the greatest impediment to the free flow and transfer of information, others being: illiteracy, censorship, volume and physical availability. It is claimed that the language barrier is even growing as more and more countries add publications in their languages[2]. Increasing amounts of primary scientific material, especially in the applied and technical sciences, are published in many non-English languages, mainly Russian, German, French and Japanese[3-8]. Studying scientific publications in geology, Reguant[9] claims that non-English languages dominate important subfields in this discipline.

There is broad consensus concerning the overwhelming importance of English in the world-wide transfer of scientific information, comprising between 40 to 80 per cent of the total international communication in the sciences, depending on the discipline [3, 6, 10-12]. Replacing the German language, English has become the 'lingua franca' of the scientific world [13]. Actually, some argue that the best scientific research of non-English speakers is published in English anyway [12,14], but this view is still highly disputed [6,15].

However, almost all studies addressing the language barrier problem, were conducted from an Anglo-Saxon perspective, limiting their scope to English-language sources or English-speakers, as already reviewed elsewhere[16]. Very little research has been devoted to language preference among non-English-speaking scholars, as compared to their English-speaking colleagues, and little attempt was made to apply additional measures, besides the simple measure of the percentages of each language used. As such, this article provides technical tools to study suggestion 10 in Rousseau's list of research suggestions[17], namely to study the receptivity to foreign literature in different fields and different countries.

Former proposed measures and critical analysis

Language analysis of cited references usually aims to reveal the languages upon which research in the investigated field relies, as well as the relative use of each. Providing that the sample of source articles (whose references are analysed) is not confined to a single language, but rather comprises several sub-samples published in various languages, it is possible to calculate the rate of 'language self-citation' (or 'own language preference') for each language group.

The measure of 'language self-citation' can be simply defined as the proportion of references written in the same language as the citing source. 'Language self-citation' is analogous to 'journal self-citation' and 'subject self-citation' [18-20] and its rate indicates the degree to which researchers in a certain field draw upon the literature published in their own language. A low rate of 'language self-citation' is assumed to indicate considerable use of, and dependence on, literature in other languages, while a high rate should indicate self-sufficiency, and independence of foreign language research literature. Seemingly, the closer the rate is to 100%, the smaller the reliance on foreign language materials.

However, as pointed out in earlier publications [16,21] this measure alone is inadequate and may create a distorted picture, since it does not indicate the quantity of foreign language material thus disregarded by those scholars, or its importance. Understandably, in fields in which English comprises a vast majority of the total world research output, English-writers with a high rate of 'language self-citation' lose much less information than in fields in which most research material is published in non-English languages.

The 'importance' of research material in 'foreign' languages is very difficult to assess and is highly disputed, as indicated above. The other problem, however, is easier to solve, providing one has data regarding the estimated proportion of the cumulative total published research, shared by each language in a specific subject field. The **actual** 'language self-citation' rate of a group of scholars publishing in a certain

language, should be related to an **expected** rate, which is based on the assumption that, by and large, with no language barriers and no material availability problems existing, scholars would cite publications in a certain language according to its share in the total research output of that field.

This refined measure of the relative language-self-citation (for a group of researchers in a certain field who publish in a certain language), which was first termed the 'linguistic isolation' coefficient [21] and later [16] renamed **'relative own-language preference' (ROLP) indicator**, is represented by the following ratio:

actual self-citation rate of that language

---

expected self-citation rate of that language

where the expected rate is expressed by an estimated proportion of this language's share in the total published research output cumulated hitherto in that field. A ROLP ratio equalling one, means a 'balanced' state in which the authors in that field cite publications in their own language proportionally to its share in the existing body of research literature, i.e. publication share. The higher the ratio lies above 1, the higher the degree of 'self-language preference' and disuse of a larger body of foreign language material. The smaller the ROLP-index, the greater the relative use of foreign language literature.

The relative greater availability and accessibility of publications in the own language, combined with the fact that these sources are more familiar, generally leads to ROLPs larger than one.

In a comparative study of language preference in the field of sociology, Yitzhaki [21] examined the language of references cited in regular original research articles published between 1989 and 1994 in nine different-nationality sociology journals: American, British, German and French (see Table 1). The 'relative own-language

preference' (ROLP) indicators, presented in Table 3 for each journal, were calculated according to the above-mentioned ratio formula (the actual percentages of cited languages are presented in Table 2). The 'expected self-citation rate' was calculated on the basis of two alternative estimates. Estimate a was based on a statistical analysis of the languages of the publications included in ***SOCIOFILE*** CD-ROM 1/74 - 12/95, while estimate b was based on the assumption that only 50% of the non-English items were included in that database.

These ROLP indicators revealed that the highest degree of 'own-language preference' (i.e. the strongest mother tongue bias) prevails among the Germans, followed by the French sociologists, while American and British ones display a much better balance, with coefficients slightly higher than 1, indicating the lowest own-language bias.

Calculating the ROLP indicators according to estimate b, which gives higher proportions to other languages in the existing body of research literature, at the expense of English, naturally shows slightly higher rates of relative language self-citation among the American and British sociologists and improves the German and French indicators. The latter ones drop considerably, but still remain much higher than the Anglo-Saxon figures, indicating a strong mother tongue 'bias' (or preference) 8 to 17 times higher than expected according to their publication share.

Although the ROLP indicators better reflect the differences in language preference, than do individual raw figures of language self-citation, critics have rightly pointed out a major drawback of the ROLP index, namely its size-dependency. Consider, as an example the fictitious data of Table 4.

Table 1 Nine sociology journals studied in this article and abbreviations used here

Theory, Culture and Society (UK)	TCS
American Sociological Review (USA)	ASR
Sociology (UK)	SOC
Sociological Review (UK)	SR
British Journal of Sociology (UK)	BJS
Revue Française de Sociologie (F)	RFS
L' Année Sociologique (F)	AS
Kölner Zeitschrift für Soziologie und Sozial Psychologie (G)	KZS
Sozialwissenschaftliche Information für Unterricht und Studium (G)	SI

Table 2 Percentages of cited references in nine sociology journals (1989-1994)

journal	English	German	French
TCS (UK)	89.6	4.9	5.4
ASR (US)	98.6	0	0.1
SOC (UK)	98.8	0.4	0.8
SR (UK)	99.3	0.7	0
BJS (UK)	99.5	0.5	0
RFS. (F)	35.5	0	64.5
AS (F)	29.6	1.7	67.5
KZS (G)	35.2	63.3	1.3
SI (G)	11.0	85.5	3.6

Table 3 ROLP-index and straight odds ratio for nine sociology journals

journal	ROLP	odds ratio
TCS (UK)	(a) 1.1	(a) 2.0
	(b) 1.3	(b) 3.9
ASR (US)	(a) 1.2	(a) 16.1
	(b) 1.4	(b) 32.1
SOC (UK)	(a) 1.2	(a) 18.8
	(b) 1.4	(b) 37.5
SR (UK)	(a) 1.2	(a) 32.4
	(b) 1.4	(b) 64.6
BJS (UK)	(a) 1.2	(a) 45.5
	(b) 1.4	(b) 90.7
RFS (F)	(a) 13.4	(a) 36.1
	(b) 8.0	(b) 20.6
AS (F)	(a) 14.1	(a) 41.2
	(b) 8.3	(b) 23.6
KZS (G)	(a) 21.1	(a) 55.8
	(b) 12.7	(b) 32.8
SI (G)	(a) 28.5	(a) 190.6
	(b) 17.1	(b) 112.0

Estimated proportions of these languages in sociological research are:

Estimate (a): English – 81.4%, French – 4.8%, German – 3.0%

Estimate (b): English – 68.7%, French – 8.1%, German – 5.0%

Table 4 Illustration of ROLP –index bias

	Self-citing rate	Publication share	ROLP-index
Language 1	100%	80%	1.25
Language 2	20%	5%	4.00

Although Language 1 is extremely parochial in its language use (it is not open at all to other languages) its ROLP is considerably smaller than that of Language 2. This means that the ROLP-index might work well in an unrealistic situation in which major languages have similar shares in the world publication 'pool', but not in an unequal scene which is biased in favour of the 'dominant' language (in our case, English) at the expense of less 'dominant' ones (in our case, German and French). It is clear that in the former case, since the English language is estimated to dominate at least about 70% of the 'market' (i.e. denominator = 70) the above ratio will yield a ROLP, for English writers, slightly higher than 1, never exceeding 1.5 (see e.g. Table 3). On the other hand, for German and French, and certainly for less 'dominant' languages, with an estimated publication share of less than 10%, the ROLP indicator tends to be considerably higher, due to a relatively low denominator (see e.g. Table 3).

In a recent paper by Bookstein and Yitzhaki [22], a more refined ROLP measure was proposed, free of the aforementioned bias, and based on the 'straight' Odds Ratio. The indicators it yields are presented in Table 3, next to the ROLP indicators.

The 'straight' Odds Ratio measure yields the ratio between two odds: the odds P_o that a scholar or a scientist, working in a certain field, will cite any given paper written in his own language, and the odds P_f that he will cite any given paper written in any other language (i.e. in a 'foreign' language).

Given that:

SCR = Proportion of own language references (as found empirically in a group of papers written in that language) = actual language self-citation rate.

Hence, 1-SCR = actual 'other-languages citation' rate.

α = Proportion of the target literature population written in one's own language (= estimated proportion of that language in the cumulative published research in the field under study).

Consequently, 1- α = Proportion of the target literature population written in other languages. Then, the final 'straight' Odds Ratio measure P_o / P_f is :

$$\frac{SCR/(1-SCR)}{\alpha/(1-\alpha)} \qquad (1)$$

As shown in Table 3, a comparison of these refined size-free 'own-language preference' indicators to the former ROLP ones, on the basis of estimate a, shows that:

1. The indicators, in absolute numbers, for all three groups of journals (Anglo-Saxon, German and French) are much higher. However, the increase in indicators for all but one (*Theory, Culture and Society*) of the Anglo-Saxon journals has been greater than that of the rest.

2. The two German journals still show the largest degree of own-language preference, but there is a much greater discrepancy between them (190 vs. 56), as versus the other journals. The *British Journal of Sociology* now shows a relatively higher indicator of own-language preference (45.5), ranking third in the list, fairly close to the second German (56), and followed by the two French journals which have lower own-language preference indicators (41 and 36).

3. The *American Sociological Review* and two of the British ones (*Sociology* and *Sociological Review*) which before had equal ROLP indicators, now differ in

their own-language preference (32.4, 18.8, 16.6) due to the greater sensitivity of the odds ratio measure to tiny differences (like 0.2%!) in the raw percentage figures of language self-citation.

However, using estimate b regarding the world wide distribution of publication languages in sociology results in great changes in both ranking (see Table 3) and values of own-language preference, due to the lower estimate it gives the English language. As shown in Table 3, all but one (*Theory, Culture and Society)* of the Anglo-Saxon journals rank much higher in the list, preceded only by the German *Sozialwissenschaftliche Information*, having relatively high indicators of own-language preference (90.7, 64.6, 37.5). In fact, these indicators are now twice those found according to estimate a since estimate b assumes that the ratio between English and other languages is about 2 (68.7 / 31.3), rather than 4 (81.4 / 18.6). Showing now a lower degree of own-language preference are the German *Kölner Zeitschrift für Soziologie*, the *American Sociological Review* and the two French ones, while the British *Theory, Culture and Society* still has the lowest own-language preference indicator (3.9).

Meaning, advantages and disadvantages of the 'Straight' Odds Ratio measure

One may say that the straight odds ratio measure measures the relation between a scholar's (say the average language L-speaking scientist) preference of his/her own language (relative to all other languages) and his/her language publication share in world research in that field (relative to the share of all other languages). In fact, it is the relation between two ratios: (1) the ratio of language self-citation to the citation of other languages, and (2) the ratio of that language publication share to the share of other languages. We may say that the odds ratio indicates to what extent his/her relative preference of his/her language is greater (or smaller) than the existing ratio of his/her language to other languages, as found in the cumulative body of research publications in that field.

The main advantage of the odds ratio measure is its being size-free, and thus independent of, and unaffected by, the 'raw' volume of body of research existing in a certain language.
The 'straight' odds ratio measure, however, has several apparent drawbacks:

1. The measure is always infinite in case the language cites only itself. The larger a language, the smaller the measure must be, even if only one's own language is cited. It does not draw conclusions concerning the internationalisation of a research discipline: if a research discipline in the Netherlands or Flanders uses only Dutch literature it receives a value infinite for the odds ratio measure (rightly so - isolation is there!) but if a research discipline in the USA only uses American (English) literature it also receives a value infinite for the odds ratio measure while it is less evident here that this discipline is isolated to the same extent as the previous one.
2. It is somewhat oversensitive, in certain cases, to negligible and insignificant differences in the percentages of language self-citation. In other words, minuscule differences in these raw proportions result in considerable differences in the measures of odds ratio for each of the journals concerned (e.g. the differences between the *British Journal of Sociology* and *Sociological Review* or between *American Sociological Review* and *Sociology).*
3. The measure is not normalised (yielding values between 0 and 1), as is often desirable.

New suggestions for correcting the ROLP-index: an axiomatic approach

The objective of this section is to propose additional measures, which may serve as better indicators of the relative own-language preference. Their adequacy for

measuring the relative self-citing rate will be determined by checking whether they meet a number of axioms we will set forward.

First, we will fix the notation. Consider the following citation matrix: $C = (c_{ij})_{lj}$, where c_{lj} denotes the number of citations given by language L to publications written in language J. Assume further that we consider n different languages (including L). The self-citing rate of language L is then defined as: $SCR(L) = c_{ll}/c_l$, where

$$c_l = \sum_{j=1}^{n} c_{lj} \qquad (2)$$

If the number of publications in language L is denoted as P(L) (in the population under study), and P denotes the total number of publications in the system, then the publication share of language L is given as:

$$\alpha(L) = \frac{P(L)}{P} \qquad (3)$$

As the target language L will be fixed we will denote $\alpha(L)$, simply by α. Similarly, the self-citing rate of language L will be denoted by c.

A measure f of the relative self-citing rate should satisfy the following axioms.

RSCR 1. The function f is a positive, real-valued continuous function of the variables α and c, defined on the set $( ]0,1[\times [0,1]) \cup \{(1,1)\}$.

We exclude the case $\alpha = 0$, because this means that the language L under study does not occur. If $\alpha = 1$, then naturally $c = 1$: if L is the only language, one can cite only this language. Continuity is required, because it is against the intuition that α or c values that differ only by a small amount, result in relative self-citing rates that

differ a lot. Moreover we require f to depend on relative parameters, not on absolute ones such as the total number of publications under consideration. This is because we want to obtain a relative measure.

RSCR 2a. $f(1,1) = 0$

b. $\forall\, \alpha \in]0,1[: f(\alpha,0) = 0$

Axiom RSCR 2a expresses the fact that if a country holds a monopoly position it can not be accused of parochialism. Axiom RSCR 2b states that if the language under study is never cited, while some publications in that language exist, the relative self-citation rate must be zero.

These axioms take care of the extreme cases. Now we concentrate on the more usual cases $\alpha \in]0,1[$ and $c \in]0,1[$.

RSCR 3. The function $f(\alpha,c)$ is strictly increasing in c (for fixed $\alpha \in]0,1[$).

This is evident: the larger the self-citation rate, the larger the relative self-citation rate for fixed α.

RSCR 4. The function $f(\alpha,c)$ is strictly decreasing in α (for fixed $c \in]0,1]$).
Indeed, the larger α, the less choice there is to cite a publication written in another language.

Finally, we require the following sensitivity property

RSCR 5. For every $c \in]0,1]$, the partial derivative

$$\left|\frac{\partial f(\alpha,c)}{\partial \alpha}\right| = -\frac{\partial f(\alpha,c)}{\partial \alpha} \quad (by\ RSCR\ 4)$$

must be strictly decreasing as a function of α.

The requirement that the partial derivative of f with respect to α must be decreasing means that f is required to change faster for small α than for larger ones (with c fixed).

Some consequences of these axioms

1. $\lim_{\alpha \to 1} f(\alpha,1) = 0$

This property follows from the continuity of f (RSCR 1) and the fact that f(1,1) = 0 (RSCR 2a)

2. For c = 1 the function f still depends on α.
This follows from RSCR 4. So RSCR 4 eliminates the odds ratio as an acceptable measure for relative self-citation preference.

3. If α tends to zero, f(α,c) tends to its maximal value (c fixed).
This follows from RSCR 4.

We will next see which measures satisfy these five axioms.

1°) The ROLP-index, c/α, clearly meets axioms 1, 2b, 3 and 4. It also satisfies RSCR 5 as

$$\left|\frac{\partial\left(\frac{c}{\alpha}\right)}{\partial\alpha}\right| = \frac{c}{\alpha^2}$$

strictly decreases in the variable α.

This shows that the ROLP-index is a good measure that only fails at axiom RSCR 2a.

2°) The odds-ratio $(c(1-\alpha)/(1-c)\alpha)$ meets axiom RSCR 1, but is does not satisfy RSCR 2a. More precisely, the odds-ratio is not defined in (1,1), but defining it to be 0 in (1,1) makes it discontinuous in (1,1) as the limit for $(\alpha,c) \rightarrow (1,1)$ where (α,c) tends to (1,1) along the diagonal $(\alpha = c)$, is equal to 1. Axiom RSCR 2b is satisfied, as are axioms RSCR 3 and 4 for $c \neq 1$. Axiom RSCR 4 is not satisfied for $c = 1$. Finally, as so RSCR 5,

$$\left|\frac{\partial}{\partial\alpha}\left(\frac{c(1-\alpha)}{\alpha(1-c)}\right)\right| = \frac{1}{\alpha^2(1-c)}$$

This expression decreases strictly in α except, again for $c = 1$. We conclude that the odds ratio fails either axiom RSCR 2a or axiom RSCR 1. It fails, moreover, axioms 4 and 5 for $c = 1$.

3°) The function $f(\alpha,c) = c\,(1 - \alpha)$. This function meets requirements RSCR 1, 2a, 2b,3 and 4. As to RSCR 5, we see that

$$\left|\frac{\partial c(1-\alpha)}{\partial\alpha}\right| = c$$

This is constant (for fixed c) and hence does not decrease in α.

4°) Finally, taking $f(\alpha,c) = c \ln(1/\alpha)$ yields a function satisfying all requirements. Indeed, the first four axioms are easy to check. As to RSCR 5, we see that

$$\left|\frac{\partial}{\partial\alpha}\left(c\ln(\frac{1}{\alpha})\right)\right| = \frac{c}{\alpha}$$

This expression decreases strictly in α.

5°) Considering, e.g. $f(\alpha,c) = c^2 \ln(1/\alpha)$, leads to another function satisfying all requirements.

We conclude that the functions

$$RSCR_1(\alpha,c) = c\ln\left(\frac{1}{\alpha}\right) \qquad (4)$$

$$RSCR_2(\alpha,c) = c^2\ln\left(\frac{1}{\alpha}\right) \qquad (5)$$

are the best ones (among those considered here) to measure the relative self-citing rate.

Further comments

1 - Interpreting the publication share α as a stimulus to the publication-citation system under study, and the relative self-citation rate as the subjective reaction or sensation of the system, then formulae (4) and (5) express a kind of Weber-Fechner relation, in the sense that the sensation depends logarithmically on the stimulus [23].

This is what we intuitively expect: the larger α, the less important the changes in the relative self-citation rate. This can also be considered as an expression of the law of diminishing returns.

2 - The functions $RSCR_1(\alpha,c)$ and $RSCR_2(\alpha,c)$ as given by formulas (4) and (5) are certainly not unique in satisfying the axioms mentioned before. Indeed, any function of the form $f(\alpha,c) = g(c)\ h(\ln(1/\alpha))$, where g and h are positive, continuous and increasing functions such that $g(0) = h(0) = 0$ and such that h' is strictly increasing satisfies all axioms.

3 - If α stays fixed, i.e. within one language, the self-citation rate determines the rank of the relative self-citing rate of the journals, people or any other subject of study.

4 – If $f(\alpha,c)$ is a function satisfying all axioms, and if m(t) is a strictly positive, continuous, increasing function, such that $m(0) = 0$, then $m(f(\alpha,c))$ also satisfies all axioms.

5 - If a measure f yields values between 0 and $+\infty$, and if one prefers a measure that takes values between 0 and 1, one can use the normalised function

$$\frac{2}{\pi} arctg(f(\alpha,c)) \qquad (6)$$

This normalisation is allowed by the previous remark. It seems, moreover, advisable to apply it.

The values of $RSCR_i(\alpha,c)$, $i = 1,2$ and of their normalised forms for the hypothetical example of Table 4 are:

$RSCR_1(L1) = RSCR_2(L1) = 0.223$, which becomes, after normalisation: 0.140
$RSCR_1(L2) = 0.599$, or 0.344 (normalised), while $RSCR_2(L2) = 0.120$ or 0.076 (normalised).

Although both RSCR-functions satisfy all the axioms we have required, this example illustrates the fact that it is still possible to obtain considerably different rankings. Based on this example our preference goes to the second one.

The values for $RSCR_i(\alpha,c)$, i = 1,2 and of their normalised forms are presented in Table 5 using both estimates (a) and (b) of the α-value of the language in question.

A comparison of the various values yielded by both $RSCR_1$ and $RSCR_2$ measures, shown in Table 5, leads to the following conclusions:

1. Ranking. Both measures yield an identical ranking of the nine journals: the British *Theory, Culture and Society* tops the list. It has the lowest RSCR values, indicating the lowest degree of 'own-language preference'. Next come the four other English-language journals, with very tiny, almost negligible, differences between each other. All four measures (normalised or not) indicate a considerable discrepancy between the English language journals and the other ones. The two French journals rank between the English journals and the German ones. Note though *that Kölner Zeitschrift für Soziologie* displays values fairly close to the French *L' Année Sociologique.*

Table 5 RSCR and normalised RSCR values for nine sociology journals

journal	$RSCR_1$	normalised $RSCR_1$	$RSCR_2$	normalised $RSCR_2$
TCS (UK)	(a) 0.184 (b) 0.336	(a) 0.116 (b) 0.207	(a) 0.165 (b) 0.301	(a) 0.104 (b) 0.186
ASR (US)	(a) 0.203 (b) 0.370	(a) 0.127 (b) 0.226	(a) 0.200 (b) 0.365	(a) 0.126 (b) 0.223
SOC (UK)	(a) 0.203 (b) 0.371	(a) 0.128 (b) 0.226	(a) 0.201 (b) 0.366	(a) 0.126 (b) 0.224
SR (UK)	(a) 0.204 (b) 0.373	(a) 0.128 (b) 0.227	(a) 0.203 (b) 0.370	(a) 0.127 (b) 0.226
BJS (UK)	(a) 0.205 (b) 0.374	(a) 0.129 (b) 0.228	(a) 0.204 (b) 0.372	(a) 0.128 (b) 0.227
RFS (F)	(a) 1.959 (b) 1.621	(a) 0.699 (b) 0.637	(a) 1.263 (b) 1.046	(a) 0.574 (b) 0.514
AS (F)	(a) 2.050 (b) 1.696	(a) 0.711 (b) 0.661	(a) 1.384 (b) 1.145	(a) 0.602 (b) 0.543
KZS (G)	(a) 2.220 (b) 1.896	(a) 0.731 (b) 0.691	(a) 1.405 (b) 1.200	(a) 0.606 (b) 0.558
SI (G)	(a) 2.998 (b) 2.561	(a) 0.795 (b) 0.763	(a) 2.563 (b) 2.190	(a) 0.763 (b) 0.727

2. Normalisation. Normalisation clearly decreases differences yielded by the 'raw' RSCR measures. The ratio between the values of *Revue Française de Sociologie* and the four English-language journals is about 10 to 1 according to the raw $RSCR_1$ measure, but it decreases to only 5.4 to 1 after normalisation. Similarly, the difference between the values of the *Kölner Zeitschrift fur Soziologie* and the French *L' Année Sociologique* decreases from 8 to 3 %, and that between the

Sozialwissenschaftliche Information and the *Kölner Zeitschrift für Soz*iologie decreases from 35% to 9%.

3. Comparing $RSCR_1$ to $RSCR_2$ values. It is evident that the latter one tends to be more size-free than the former. It is probably less affected by the greater publication share of the dominant language (English), yielding lower RSCR values for the other languages and thus reducing the differences between the English-language journals and the French and German ones. In other words, giving almost identical values (to those of $RSCR_1$) for the English-language journals and lower values (than $RSCR_1$) for the French and German ones (independent of estimate a or b). We can say that $RSCR_2$ reduces gaps between the two groups.

4. Effects of a and b estimates. As expected, using the b estimates favouring the non-English languages, increases on the one hand the RSCR values ('raw' as well as normalised) for all English-language journals, and on the other hand, decreases the RSCR-values for all non-English journals, resulting in narrowing the gap between these two groups.

5. Group comparison. While the RSCR values for all English-language journals fluctuate between 0.13 and 0.23 (depending on the estimate) all non-English ones display values higher than 0.5 (even according to estimate b and $RSCR_2$). Thus one may say that the basic findings and conclusions reached in earlier articles, namely that the four non-English sociology journals display in their references a relatively higher degree of own-language preference, still holds. Using the refined measures proposed here only reduces the gap between these two groups.

We finally note that everything that has been stated for the relative self-citation rate can also be developed for the relative self-cited rate. The issues studied here deserve, in our opinion, a more formal approach. Moreover, they should be generalised in order to study the openness of one language with respect to another one. This will be done in forthcoming work [24] .

Conclusion

The new refined RSCR measures, proposed in this article, still give the nine studied sociology journals the same rank as former measures mentioned in the literature, although the gaps between individual journals, and especially between the two groups (English-language journals and non-English ones) narrows. The new measures have clear advantages lacking in the former ones, and thus serve better the purpose of measuring 'relative own-language preference'.

Acknowledgement. The authors thank Prof. A. Bookstein for helpful e-mail and fax exchanges concerning the topics discussed in this article.

References

1. J.A. LARGE, *The Foreign-Language Barrier: Problems in Scientific Communication,* André Deutch, London, 1983.
2. P. MAYORCAS-COHEN, Review of *The Foreign-Language Barrier: Problems in Scientific Communication* by J.A. Large, *Journal of Documentation,* 41 (1985) 56-58.
3. D.N. WOOD, The foreign-language problem facing scientists and technologists in the U.K.- report of a recent survey, *Journal of Documentation* 23(2) (1967) 117-130.
4. H. WELLISH, Linguistic and semantic problems in the use of English-language information services in non-English-speaking countries, *International Library Review* 5 (1973) 147- 162.
5. M. JAGODZINSKI-SIGOGNEAU, J.P. COURTIAL, B. LATOUR, How to measure the degree of independence of a research system? *Scientometrics* 4 (1982) 119-133.

6. J. MICHEL, Linguistic and political barriers in the international transfer of information in science and technology, *Journal of Information Science* 5 (1982) 131-135.
7. D. BEADELY, European elites envy American cohesion, *Science* 260 (1993) 1738-1739.
8. S. REGUANT, What lingua franca? *Nature* 361 (1993) 107.
9. S. REGUANT, English as lingua franca in geological scientific research; a bibliometric study, *Scientometrics* 29 (1994) 335-351.
10. J.E. HOLMSTROM, The foreign language barrier, In: J. SHERROD, A. HODINA (Eds) *Reader in Science Information*, Washington, D.C., Microcard, 1973, 94-103.
11. H. ENGELBERT, The terminological barrier to the understanding of scientific and technological literature, *Translation News* 6(2) (1976) 2-10.
12. C.R.H. INMAN, Scientific Publications in English and developing countries: a report of a survey of scientists' experiences, *Journal of Information Science* 6 (1983) 159-164.
13. R.G. THORP, H. SCHUR, D. BAWDEN, J.R. JOICE, The foreign language barrier: a study among pharmaceutical research workers, *Journal of Information Science* 14 (1988) 17-24.
14. E. GARFIELD, Do French scientists who publish outside of France and/or in English do better research? In: E. GARFIELD (ed.) *Essays of an Information Scientist*. vol. 3. Philadelphia, ISI Press, 1980, 498-503.
15. G.K.L. CHAN, The foreign language barrier in science and technology. *International Library Review* 8 (1976) 317-325.
16. M. YITZHAKI, The 'language preference' in sociology: measures of 'language self-citation', 'relative own-language preference indicator', and 'mutual use of languages', *Scientometrics* 41 (1998) 243-254.

17.R. ROUSSEAU, Suggestions for research topics in citation and publication analysis, *Library Science with a Slant to Documentation and Information Studies* 32 (1995) 3-12.

18.P. EARLE, B. VICKERY, Social science research literature use in the UK as indicated by citations, *Journal of Documentation* 25 (1969) 123-141.

19.D. NICHOLAS, M. RITCHIE, *Literature and Bibliometrics*, Clive Bingley, London, 1978.

20.M. YITZHAKI, Subject dispersion and discipline affiliation. *Scripta Hierosolymitana* (Publications of the Hebrew University of Jerusalem) 29 (1989) 217-249.

21.M. YITZHAKI, The language barrier in the humanities: the case of biblical studies, In: L. EGGHE, R. ROUSSEAU (Eds) *Informetrics 87/88; Selected Proceedings of the First International Conference on Bibliometrics, Diepenbeek, Belgium, Aug. 1987* Elsevier, Amsterdam, 1988, 301-314.

22. A. BOOKSTEIN, M. YITZHAKI, Own-language preference: a new measure of 'relative language self-citation', In: *Proceedings of the Seventh Conference of the International Society for Scientometrics and Informetrics, Colima (Mexico) 1999* (to appear).

23. F.S. ROBERTS, *Measurement Theory,* Addison-Wesley, London, 1979.

24. L. EGGHE, R. ROUSSEAU, Partial orders and measures for language preferences, preprint, 1999.